PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/15, A61K 37/64
C07K 13/00, C12P 21/02

A1

(11) International Publication Number: WO 91/03557

(43) International Publication Date: 21 March 1991 (21.03.91)

(21) International Application Number: PCT/DK90/00225

(22) International Filing Date: 29 August 1990 (29.08.90)

(30) Priority data:

4235/89	29 August 1989 (29.08.89)	DK
4236/89	29 August 1989 (29.08.89)	DK
4237/89	29 August 1989 (29.08.89)	DK

(71) Applicant *(for all designated States except US)*: NOVO NORDISK A/S [DK/DK]; Novo Alle, DK-2880 Bagsværd (DK).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : ESPER, Boel [DK/DK]; Lyngbakkevej 25, Søllerød, DK-2840 Holte (DK). LARS, Sottrup-Jensen [DK/DK]; Johannes Evaldsvej 75, DK-8230 Åbyhøj (DK).

(74) Common Representative: NOVO NORDISK A/S; Patent Department, Novo Alle, DK-2880 Bagsvaerd (DK).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent)*, DK (European patent), ES (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.

Published
With international search report.

(54) Title: EXPRESSION OF ALPHA-MACROGLOBULINS

(57) Abstract

α-Macroglobulins, especially human α_2-macroglobulin, variants, fragments or derivatives thereof is produced by recombinant technology. The products are useful as additives to growth media, as proteinase inhibitors, as carrier in enzyme replacement therapy, and as DNA carrier in gene therapy.

* See back of page

DESIGNATIONS OF "DE"

Until further notice, any designation of "DE" in any international application whose international filing date is prior to October 3, 1990, shall have effect in the territory of the Federal Republic of Germany with the exception of the territory of the former German Democratic Republic.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**ES**	Spain	**MC**	Monaco
AU	Australia	**FI**	Finland	**MG**	Madagascar
BB	Barbados	**FR**	France	**ML**	Mali
BE	Belgium	**GA**	Gabon	**MR**	Mauritania
BF	Burkina Fasso	**GB**	United Kingdom	**MW**	Malawi
BG	Bulgaria	**GR**	Greece	**NL**	Netherlands
BJ	Benin	**HU**	Hungary	**NO**	Norway
BR	Brazil	**IT**	Italy	**PL**	Poland
CA	Canada	**JP**	Japan	**RO**	Romania
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SN**	Senegal
CM	Cameroon	**LI**	Liechtenstein	**SU**	Soviet Union
DE	Germany	**LK**	Sri Lanka	**TD**	Chad
DK	Denmark	**LU**	Luxembourg	**TG**	Togo
				US	United States of America

Title: Expression of alpha-macroglobulins

FIELD OF THE INVENTION

The present invention relates to the expression of α-macroglobulins, derivatives and variants thereof, and especially the expression of the human α_2-macroglobulin (α_2M) in an active form in mammalian cells, and the expression of genetically engineered variants thereof. The use of such recombinant α-macroglobulins, especially recombinant α_2M(rα_2M) and variants is described with examples from the fields of medicine for therapeutic purposes, and the development of novel defined growth media for propagation of mammalian cells in culture.

BACKGROUND OF THE INVENTION.

BIOCHEMISTRY OF α_2-MACROGLOBULIN (α_2M).

The proteinase binding glycoprotein α_2M, which is synthesized in the liver, constitute together with the complement proteins C3, C4 and C5 a separate class of structurally and functionally related large plasma proteins. For a recent review see (Sottrup-Jensen, L. (1987) in: The Plasma Proteins (Putnam, F.W., ed.) 2nd Ed., 5: 191-291, Academic Press, Orlando, FL).

Apart from C5 these proteins contain an internal ß-cysteinyl-γ-glutamyl thiol ester, which enables the proteolytically activated forms of α_2M, C3, and C4 to participate in characteristic covalent binding reactions (Sottrup-Jensen, L., et al., (1980) FEBS Lett. 121: 275-280; Salvesen, G.S. and Barrett, A.J., (1981) Biochem. J. 187: 695-701). The thiol ester structure, which in the active proteins can be slowly cleaved by a number of small nitrogen nucleophiles, constitutes a unique type of postsynthetic modification of proteins, and plays a prominent role in the biological properties of α_2M. The presence of the active thiol esters in α_2M is revealed by a characteristic pattern of heat fragmentation (Harpel, P.C., et al., (1979) J. Biol. Chem. 254: 8869-8878).

Traditionally, α_2M has been studied within the context of plasma proteinase inhibitors, although by several criteria it is unique. Whereas most plasma proteinase inhibitors are monomeric proteins of roughly similar size, containing approximately 430-500 residues, α_2M is a tetramer whose 180-kD subunits contain 1451 residues (Sottrup-Jensen et al., (1984) J. Biol. Chem. 259: 8318-8327).

Furthermore, in contrast to most other proteinase inhibitors, which form 1:1 complexes with serine proteinases engaging the active site

of the proteinase and the reactive site of the inhibitor, α_2M forms complexes with a broad spectrum of proteinases differing in their substrate specificity and catalytic mechanism e.g.: trypsin, leucocyte elastase, chymotrypsin, pancreatic elastase, cathepsin G, plasmin, plasma kallikrein and thrombin.

The second-order rate constant for association between these proteinases and α_2M varies by several orders of magnitude. Both 1:1 and 2:1 proteinase-α_2M complexes can be formed, and the disulfide-bridged dimer (360 kD) appears to be the functional unit of α_2M (Sottrup-Jensen, L. (1987) in: The Plasma Proteins (Putnam, F.W., ed.) 2nd Ed., 5: 191-291, Academic Press, Orlando, FL). Contrary to "classical" proteinase inhibitor complexes the α_2M bound proteinase is still active, especially toward small synthetic substrates (Sottrup-Jensen, L. (1987) in: "The Plasma Proteins" (Putnam, F.W., ed.) 2nd Ed., 5: 191-291, Academic Press, Orlando, FL).

The mechanism of proteinase binding by α_2M has been described by the "trap" (Barrett, A.J. and Starkey, P.M. (1973) Biochem. J. 133: 709-724), where proteolytic cleavage of a particularly exposed peptide stretch near the middle of the 180-kD subunit (the "bait" region) results in a conformational change of the α_2M tetramer, thereby entrapping the proteinase. The nature of the essentially irreversible proteinase complex formation with α_2M has long remained elusive. However, recent investigations show that a major fraction (typically > 80-90 % of the trapped proteinase is also covalently bound through epsilon-lysyl(proteinase)-γ-glutamyl(α_2M) bonds (Sottrup-Jensen, L. et al., (1981) FEBS Lett. 128: 127-132; Sand, O. et al., (1985) J. Biol. Chem. 260: 15723-15735; Pochon, F. et al., (1987) FEBS Lett. 217: 101-105).

PHYSIOLOGICAL ASPECTS OF PROTEINASE-α_2M INTERACTIONS.

Since the α_2M-proteinase complexes are rapidly cleared from the circulation (Ohlsson, K. (1971) Acta Physiol. Scand. 81: 269-272; Imber, M.J. and Pizzo, S.V. (1981) J. Biol. Chem. 256: 8134-8139.) a general role as a "clearing vehicle" for plasma proteinases has been envisaged.

The main physiological targets may include proteinases of the coagulation and fibrinolysis systems and plasma kallikrein, and perhaps also proteinases like leucocyte elastase, cathepsin G and collagenases and other proteinases released during cellular turnover (Sottrup-Jensen, L. and Birkedal-Hansen, H. (1989) J. Biol. Chem. 264: 393-401).

Although α_2M may be largely confined to the vasculature in healthy uninflamed tissues, the inhibitor and its proteinase complexes are found at near plasma levels in inflammatory exudates of rheumatoid joints and gingival

crevicular fluids (Tollefsen, T. and Saltved, E. (1980) J. Periodont. Res. 15: 96-106; Borth, W., et al., (1983) Ann. N. Y. Acad. Sci. 421: 377-381).

While plasma α_2M appear to be synthesized in the liver (Schreiber, G. (1987) in: "The Plasma Proteins" (Putnam, F.W., ed) 2nd Ed., 5: 294-363, Academic Press, Orlando, FL.) other sites of synthesis exist. Several cell strains in culture have been shown to produce α_2M including fibroblasts (Mosher, D.F., et al., (1977) J. Clin. Invest. 60: 1036-1045) and monocytes-/macrophages (Hovi, T., et al., (1977) J. Exp. Med. 145: 1580-1589).

Whereas hepatocytes and Kupffer cells of the liver are most important for clearance of α_2M-proteinase complexes in plasma (Davidsen, O., et al., (1985) Biochim. Biophys. Acta 846: 85-92), fibroblasts (Van Leuven, F., et al., (1979) J. Biol. Chem. 254: 5155-5160; Mosher, D.F. and Vaheri, A. (1980) Biochim. Biophys. Acta 627: 113-122) and macrophages (Debanne, M.T., et al., (1975) Biochim. Biophys. Acta 411: 295-304; Kaplan, J. and Nielsen, M.L. (1979) J. Biol. Chem. 254: 7323-7328) also possess receptors for α_2M-proteinase complexes.

These observations suggest that there may be a considerable extravascular turnover of α_2M perhaps primarily carrying proteinases functioning in the cellular micro environment (Sottrup-Jensen, L. and Birkedal-Hansen, H. (1989) J. Biol. Chem. 264: 393-401).

SUMMARY OF THE INVENTION

Briefly stated, the present invention discloses a method for the production of recombinant α-macroglobulins, and especially human α_2M, and variants thereof in an active form.

Within a preferred embodiment, the cultured host cell is an eukaryotic cell such as a mammalian cell or cells derived from organisms such as insects, plants, yeast or other fungi, such as Aspergillus.

The invention further relates to DNA sequences comprising a gene encoding for the expression of human α_2M and variants thereof, vectors comprising such DNA sequences, and suitable hosts transformed with such vectors.

Yet another aspect of the invention is the use of recombinant α_2M and variants thereof as a protein carrier in enzyme replacement therapy (ERT).

Yet another aspect of the invention is the use of recombinant α_2M and variants thereof as a DNA carrier in gene therapy.

Further aspects of the invention relates to the use of recombinant α-macroglobulins, especially human α_2M, and variants thereof as

constituents of growth media, either as an additive or co-expressed with a desired gene product.

DEFINITIONS

Prior to setting forth the invention it may be helpful for an understanding thereof to set forth definitions of certain terms to be used hereafter.

Complementary DNA or cDNA: A DNA molecule or sequence which have been enzymatically synthesized from sequences present in a mRNA template.

DNA Construct: A DNA molecule, or a clone of such a molecule, either single- or double-stranded, which may be isolated in partial form from a naturally occurring gene or which has been modified to contain segments of DNA which are combined and juxtaposed in a manner which would not otherwise exist in nature.

Plasmid or Vector: A DNA construct containing genetic information which may provide for its replication when inserted into a host cell. A plasmid generally contains at least one gene sequence to be expressed in the host cell, as well as sequences encoding functions which facilitate such gene expression, including promoters and transcription initiation sites. It may be a linear or closed circular molecule.

Joined: DNA sequences are said to be joined when the 5′ and 3′ ends of one sequence are attached by phosphodiester bonds to the 3′ and 5′ ends, respectively, of an adjacent sequence. Joining may be achieved by such methods as ligation of blunt or cohesive termini, by synthesis of joined sequences through cDNA cloning, or by removal of intervening sequences through a process of directed mutagenesis.

Variant: A peptide related to the original peptide, but wherein the amino acid sequence has been altered through mutation of the gene encoding the original peptide.

ABBREVIATIONS

AMINO ACIDS

A	=	Ala	=	Alanine
V	=	Val	=	Valine
L	=	Leu	=	Leucine
I	=	Ile	=	Isoleucine
P	=	Pro	=	Proline
F	=	Phe	=	Phenylalanine
W	=	Trp	=	Tryptophan
M	=	Met	=	Methionine
G	=	Gly	=	Glycine
S	=	Ser	=	Serine
T	=	Thr	=	Threonine
C	=	Cys	=	Cysteine
Y	=	Tyr	=	Tyrosine
N	=	Asn	=	Asparagine
Q	=	Gln	=	Glutamine
D	=	Asp	=	Aspartic Acid
E	=	Glu	=	Glutamic Acid
K	=	Lys	=	Lysine
R	=	Arg	=	Arginine
H	=	His	=	Histidine

NUCLEIC ACID BASES

A	=	Adenine
G	=	Guanine
C	=	Cytosine
T	=	Thymine(only in DNA)
U	=	Uracil (only in RNA)

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1a illustrates the construction of plasmid p1136.

Figure 1b illustrates the construction of plasmid p1167.

Figure 2 illustrates the structure of plasmid p1167.

Figure 3 illustrates a gel electrophoresis (10 - 20 % SDS-PAGE) of the thermal fragmentation products generated from α_2M and rα_2M.

Figure 4 illustrates a gel electrophoresis of the thermal fragmentation products generated from methylamine treated α_2M and rα_2M.

Figure 5 illustrates a gel electrophoresis (SDS-PAGE) of the reaction products generated from trypsin treatment of α_2M and rα_2M.

Figure 6 illustrates a gel electrophoresis of the reaction products generated from trypsin treatment of methylamine-treated α_2M and rα_2M.

Figure 7 illustrates a "rate gel" electrophoresis of unreacted native -and trypsin treated α_2M and rα_2M.

Figure 8 illustrates a "rate gel" electrophoresis of unreacted native -and methylamine treated α_2M and rα_2M.

Figure 9 illustrates the chromatograms of α_2M and rα_2M on a Superose 6 column.

Figure 10 illustrates the gel electrophoresis (10 - 20 % reducing SDS-PAGE) of the reaction products from chymotrypsin treated human α_2M, human PZP and rα_2M-PZP.

Figure 11 illustrates the gel electrophoresis (10 - 20 % reducing SDS-PAGE) of the reaction products from elastase treated human α_2M, human PZP and rα_2M-PZP.

Figure 12 illustrates the gel electrophoresis (10 - 20 % reducing SDS-PAGE) of the reaction products from trypsin treated human α_2M, human PZP and rα_2M-PZP.

Figure 13 illustrates the gel electrophoresis (10 - 20 % reducing SDS-PAGE) of the reaction products from Staphylococcus aureus Glu-specific protease treated human α_2M, human PZP and rα_2M-PZP.

DETAILED DESCRIPTION OF THE INVENTION

According to the invention there is provided a process for the production of α-macroglobulins, especially human α_2-macroglobulin, or fragments or derivatives, including variants thereof, wherein a functionally operative expression vector comprising a gene encoding for the expression of a α-macroglobulin, especially human α_2-macroglobulin, or fragments or derivatives thereof, including variants, or alleles of such a gene, is introduced into a suitable host capable of expressing said gene, said host is cultured in a suitable nutrient medium containing sources of assimilable carbon and nitrogen and other essential nutrients, and the expressed α-macroglobulin, especially human α_2-macroglobulin, or fragments or derivatives thereof is recovered.

Many proteins synthesized particularly in mammalian cells undergo post-translational modification (processing) of one kind or the other.

Depending on the final destination and on the specific function of a newly synthesized protein, it may go through a number of processing steps leading to covalent modifications such as e.g.: glycosylation, γ-carboxylation, ß-hydroxylation, sulphatation, amidation, thiol ester formation, phosphorylation, proteolytic cleavage at precursor processing sites, fatty acylation (Rosner, M.R. (1986). in: "Mammalian Cell Technology", (Thilly, W.G. ed), Butterworth Publishers, Stoneham, MA.: 63-89).

Proteins of various sizes and with a variety of different post-translational modifications have been successfully expressed in transformed heterologous mammalian host cells using recombinant DNA technology. A few examples: Human coagulation factors VIIa and IX have been expressed in transformed BHK (Syrian Baby Hamster Kidney) cells with correct post-translational modifications such as γ-carboxylation and glycosylation (Thim, L. et al., (1988) Biochemistry 27: 7785-7793; Busby, S. et al., (1985) Nature 316: 271-273). Human Platelet-derived Growth Factor AB heterodimer has been expressed in transformed CHO (Chinese Hamster Ovary) cells with correct processing of the A and B chain precursors and correct assembly of the AB heterodimer. Human coagulation factor VIII has been expressed in transformed CHO cells with correct processing of the precursor leading to a two chain molecule that can be activated by thrombin and factor Xa (Kaufman, R.J. et al., (1988) J. Biol. Chem. 263: 6352-6362; Pittman, D.D. and Kaufman, R.J. (1988) Proc. Natl. Acad. Sci. USA 85: 2429-2433).

So far, there have been no reports on the heterologous expression of proteins in which the formation of an active thiol ester is a prominent post-translational modification.

The biosynthesis of the internal thiol ester in the third component (C3) of complement from rabbit has been investigated (Iijima, M. et al., (1984) J. Biochem. 96: 1539-1546). Rabbit liver mRNA was translated *in vitro* in a rabbit reticulocyte lysate system, and the synthesized C3 specific products did not incorporate radio labelled methylamine. On the other hand radio labelled iodoacetamide reacted with the synthesized C3 specific products; these results indicated the presence in the primary C3 specific translation product of a free thiol group instead of a reactive thiol ester. If a liver homogenate supernatant (S-13) including cytosol and microsomes was included, the C3 specific product could now incorporate methylamine. By increasing the concentration of the S-13 component(s), the incorporation of methylamine in C3 specific products was increased, and at the same time incorporation of iodoacetamide decreased. If the S-13 fraction was treated at 65°C for 5 min, the activity was completely lost.

The results from this investigation strongly suggest an involvement of a transglutaminase-like or other type of enzyme in the posttranslational formation of an active thiol ester in rabbit C3. There are no similar investigations addressing the formation of the thiol ester in other α-macroglobulins, e.g. α_2M, but from analogy and homology considerations, it is expected that a similar mechanism is responsible for the formation of thiol esters in other α-macroglobulins synthesized in the mammalian liver.

Through this investigation a number of developments were done which also are deemed to be encompassed of the present invention. These include DNA sequences comprising a gene encoding for the expression of α-macroglobulins, especially human α_2-macroglobulin, or fragments or derivatives and variants thereof as exemplified in SEQ ID NO:1 and SEQ ID NO:3.

Another aspect of the invention relates to functionally operative expression vectors comprising a gene encoding for the expression of at least one α-macroglobulin, especially human α_2-macroglobulin or fragments or derivatives and variants thereof, or alleles of such a gene.

Such vectors preferably further comprise regulatory elements necessary for the stable maintenance of said vector in mammalian cells.

Also, such vectors may further include sequences providing for the processing and secretion of the expressed product.

In relation to the use of recombinant α-macroglobulins, and especially rα_2M, in growth media it may be co-expressed with another desired gene product, and consequently the vectors of the invention may further comprise one or more other genes encoding for a desired gene product.

The invention further relates to transformed hosts comprising a functionally operative expression vector according to the invention comprising a gene encoding for the expression of human α_2-macroglobulin or fragments or derivatives and variants thereof, or alleles of such a gene.

The host may be selected from the group comprising a bacterial strain, a fungal strain, a mammalian cell line, or a mammal, especially a fungus, such as belonging to the genus Aspergillus, or a yeast strain, preferably belonging to the genus Saccharomyces.

Another preferred type of host is a mammalian cell line, preferably a Syrian Baby Hamster Kidney (BHK) cell line, and especially the one which is available from ATCC under No. CRL 1632.

The invention further relates to the recombinant human α_2-macroglobulin or a variant thereof in an active form having the amino acid sequence of SEQ ID NO:2, or SEQ ID NO:4.

APPLICATIONS OF α-MACROGLOBULINS, ESPECIALLY rα_2M.

The present invention discloses applications of α-macroglobulins, and especially rα_2M. These should be regarded not as limitations but as a few examples among many for the use of recombinant derived α-macroglobulins.

α-MACROGLOBULINS AS CONSTITUENTS OF DEFINED GROWTH MEDIA.

Degradation of specific heterologous products produced in either transformed or non-transformed mammalian cells is a potential problem in the production of recombinant products. This is due to the fact that many host cells secretes one or more different proteinases.

When a production cell line is grown in the presence of e.g. 10 % fetal calf serum, such proteolytic degradation of secreted recombinant or native protein products is a minor problem due to a buffering effect of the added serum proteins.

However, the use of fetal calf serum in the large scale growth (fermentation) of mammalian production cell lines is not a desirable situation for a number of reasons. First of all fetal calf serum is a very costly constituent of complex growth media; second, the demand for fetal calf serum from a growing biopharmaceutical industry might not be easily fulfilled in the future, and third, the use of fetal calf serum constitutes a potential quality control problem in the production of pharmaceuticals intended for use in humans.

To circumvent these problems, efforts can be expected in the field of development of defined growth media for use with mammalian cells.

Addition of various proteinase inhibitors to such new defined growth media will be required to ensure the integrity of the secreted products. Alternatively, the producer cell line might, through genetic engineering, be endowed with the capacity to produce and secrete proteinase inhibitors along with the desired product(s).

α-Macroglobulins, and especially Human α_2M, are proteinase inhibitors of broad specificity, and they are therefore according to the invention used as constituents of defined growth media for mammalian cells, either as a medium additive or as a product co-produced with the desired product.

The target sites for a number of different proteinases, e.g. bovine trypsin, Streptomyces griseus trypsin, papain, porcine elastase, bovine chymosin, bovine chymotrypsin, Staphylococcus aureus strain V8 proteinase, human plasmin, bovine thrombin, thermolysin, subtilisin Novo and Streptomyces griseus proteinase B have been mapped in the bait region of human α_2M (Mortensen, S.B., et al., (1981) FEBS Lett. 135: 295-300) and other α-macroglobulins (Sottrup-Jensen, L., Sand, O., Kristensen, L. and Fey, G.H. J.Biol.Chem. 264,15781-15789, 1989). It is evident that α_2M and the other α-macroglobulins as proteinase inhibitors have broad specificities.

In those situations, where the proteinase inhibitory spectrum of a α-macroglobulin, such as α_2M, is not sufficient for the prevention of product degradation, it is possible through site specific mutation, protein engineering, etc. to change the proteinase inhibitor specificity of the α-macroglobulin, such as α_2M. Incorporation of desirable specific proteinase target sites in the bait region of recombinant α_2M will change the inhibitor specificity of the mutated α_2M. Furthermore it is possible through genetic engineering to construct novel specific or general proteinase target sites in the bait region of a α-macroglobulin in order to enhance its versatility as a proteinase inhibitor of specific or broad inhibitory spectrum. Furthermore it is possible to remove specific target sites in an α-macroglobulin in order to avoid degradation of the variant in question by certain proteases in the circulation that will already be inhibited through the action of naturally present proteinase inhibitors.

The production of recombinant products in fungi, such as species and strains of e.g. Aspergillus and Saccharomyces also meets with potential problems of product degradation. In some cases it is possible to isolate proteinase negative mutants of desirable production strains. This might not always be the case, and co-expression of α-macroglobulins, such as α_2M or α_2M-mutants together with a desirable product may inhibit proteolysis of the product in question.

α-MACROGLOBULIN MUTANTS AS SPECIFIC PROTEINASE INHIBITORS.

The amino acid sequence of the bait region of α-macroglobulins defines the specificity of the α-macroglobulin towards different proteinases. A comparison of cleavage patterns for different proteinases and bait region sequences in five mammalian α-macroglobulins has recently been published (Sottrup-Jensen, L., Sand, O., Kristensen, L. and Fey, G.H. The α-macroglobulin bait region. Sequence diversity and localization of cleavage sites for proteinases in five mammalian α-macroglobulins. J. Biol. Chem. 264,

15781-15789, 1989). It has previously been clearly demonstrated that the bait region in each species of α-macroglobulin is the major determinant of proteinase inhibitor specificity. The present invention demonstrates the possibility of modulating the inhibitor specificity of human α_2M by alterations of proteinase target sites in the bait region.

In the present invention it is demonstrated that the bait region of human α_2M (residues 690 to 730 in SEQ ID NO:2) can be mutated at will to obtain a new proteinase inhibitor profile of this macroglobulin. The example presented in the present invention describes the construction of a hybrid macroglobulin. In this hybrid the bait region from human pregnancy zone protein (PZP) was introduced into human α_2M, from which the native bait region had been removed. The hybrid molecule, which was constructed by the use of recombinant DNA technology, revealed a proteinase inhibitor profile similar to the inhibitor profile of PZP.

The invention thus demonstrates the possibility to design and produce proteinase inhibitors with altered and new inhibitor specificities at will.

This finding is important for the design of new proteinase inhibitors. Due to the low antigenicity the bait region in macroglobulins (Van Leuven, F., Marynen, P., Cassiman, J.-J. and Van den Berghe, H. Mapping of structure-function relationships in proteins with a panel of monoclonal antibodies. A study on human alpha-2-macroglobulin. J. Immunol. Methods 111, 39-49, 1988, and Delain, E., Barray, M., Tapon-Bretaudiere, J., Pochon, F., Marynen, P., Cassiman, J.-J., Van den Berghe, H. and Van Leuven, F. The Molecular Organization of Human alpha2-Macroglobulin. An Immunoelectron microscopic study with monoclonal antibodies. J. Biol. Chem. 263, 2981-2989, 1988) it is now possible, by the use of the technology described in the present invention, to design non-immunogenic new proteinase inhibitors that can be used e.g. in the treatment of any disease, where aggressive proteinases constitute a threat to the health of man.

In the present specification the production of α_2M variants is described by the construction of a hybrid macroglobulin. It is clear to the skilled person in the art that changes also could be obtained through other genetic engineering methods, such as described in International Publication No. WO 89/06279 (NOVO INDUSTRI A/S). Also it is clear that other α-macroglobulins could be employed instead of the human α_2M, such as those mentioned in Sottrup-Jensen, L. et al. (1989), supra.

rα_2M AS A PROTEIN CARRIER IN ENZYME REPLACEMENT THERAPY.

A different application of α_2M is its use as a carrier of macromolecules such as proteins and nucleic acids. When α_2M reacts with and forms a complex with a proteinase in solution, α_2M may bind other proteins (also non-proteinase proteins) present in that solution (Salvesen, G.S. et al., (1981) Biochem. J. 195: 453-461). In the case of Fabry's disease, which is an X-chromosome linked disorder of glycosphingolipid metabolism, it has recently been demonstrated that α_2M can function as a carrier in an in vitro model of enzyme replacement therapy (ERT) (Osada, T., et al., (1987) Biochem. Biophys. Res. Commu. 142: 100-106). α_2M was conjugated to coffee bean α-galactosidase through the action of trypsin, and the formed complex was internalized through α_2M-receptor specific (Van Leuven, F., et al., (1981) J. Biol. Chem. 256: 9016-9022) endocytosis and delivered to the lysosomes, which is the target organelle for α_2M-receptor mediated internalization of α_2M-proteinase complexes (Willingham, M.C. and Pastan, I., (1980) Cell 21: 67-77).

Such a scheme in ERT provides a method of internalization to the lysosome of the enzyme in question and at the same time it might alleviate potential antigenicity problems arising from the use of heterologous enzymes in therapy. One limitation in this type of ERT (Osada, T., et al., (1987) Biochem. Biophys. Res. Commu. 142: 100-106) would be the types of potential target cells that could be treated by this protocol. Obviously, they would have to express the α_2M-receptor. In a future development of the system, the possibility might exist to redesign the cell specificity of α_2M internalization by exchanging the receptor binding domain of α_2M with other receptor ligands. Hereby α_2M-mutants could be designed to enter any cell type known to express a specific internalizable receptor.

This type of development would of course require a system for the production of recombinant derived α_2M. The use of native human α_2M as a carrier in ERT (as described above) is undesirable due to the now well known risks of the employment of blood derived products in the treatment of human disease.

The production of recombinant α_2M in accordance with the present invention alleviates this problem by providing for large scale production of rα_2M.

rα_2M AS A DNA CARRIER IN GENE THERAPY.

Advances in gene transfer into mammalian cells have opened for the possibility of the treatment of a number of genetic disorders through

gene therapy. A major problem in gene therapy will be the specific targeting of genes into the appropriate cells within the body. (Williamson, B., (1982) Nature 298: 416-418; Anderson, W.F., (1984) Science 226: 401-409; Parkman, R., (1986) Science 232: 1373-1378).

It was recently described that a constructed foreign gene containing the chloramphenicol acetyltransferase (CAT) on a bacterial plasmid could be targeted to the liver of rats by specific receptor directed internalization (Wu, G.Y. and Wu, C.H. (1988) J. Biol. Chem. 263: 14621-14624). The DNA carrier consisted of a galactose-terminal (asialo)glycoprotein and asialoorosomucoid covalently linked to poly-L-lysine. The polycation poly-L-lysine can bind DNA in a strong non-covalent and nondamaging interaction. It was demonstrated that complex bound DNA was internalized by cell-surface asialoglycoprotein receptors that are unique to hepatocytes. The complex was injected intravenously, and upon analysis only the liver expressed the CAT activity.

In the present invention the use of $r\alpha_2M$ as a carrier of DNA in gene therapy is suggested. Reaction of $r\alpha_2M$ with a proteinase such as trypsin or with methylamine in the presence of covalently closed circular plasmid DNA is likely to result in partial or total entrapment of DNA within the complexing α_2M molecule. After intravenous injection of such complexes with exposed receptor binding domains, the complex will be rapidly cleared from the blood and internalized in specific target cells, such as hepatocytes and Kupffer cells. Through protein engineering on the receptor binding domain of $r\alpha_2M$ it will be possible to design a DNA carrier specific for other cell types. The advantage in this system as compared to the above described system using the asialoglycoprotein receptor is, that it will not be necessary to identify different DNA carrier systems for each new cell type.

EXAMPLES

Materials and methods:

Microorganisms and cell lines

E. coli K12 (MC1061) is available from e.g. Stratagene Inc., 11099 North Torrey Pines Rd., La Jolla, California 92037.

HepG2 (Human hepatoblastoma cell line) is freely available from American Type Culture Collection, under No. HB 8065.

BHK (Syrian Hamster Kidney cell line, thymidine kinase mutant line tk^{-}s13, (Waechter and Baserga (1982) Proc. Natl. Acad. Sci. USA 79:

1106-1110); is freely available from American Type Culture Collection, under No. CRL 1632.

Plasmids and vectors

Plasmids pCDVI-PL and pSP62-K2 are available from Dr. Tasuku Honjo, Faculty of Medicine, Kyoto University, Kyoto 606, Japan. pSP62-K2 was derived from the plasmid pSP62-PL (available from New England Nuclear/Du Pont (U.K.) Ltd., Wedgwood Way, Stevenage, Hertfordshire, SG14QN) as described (Noma et al., (1986) Nature, 319: 640-646). pCDVI-PL was derived from pcDV1 (Okayama, H. and Berg, P. (1983) Molec. cell. Biol. 3: 280-289) as described (Noma et al., (1986) Nature, 319: 640-646).

M13mp18 is available from Pharmacia LKB Biotechnology (catalog # 27-1552-01) (Norrander, J., Kempe, T. and Messing, J. Gene 26: 101-106, 1983).

M13mp19 is available from e.g. International Biotechnologies, Inc., P.O. Box 9558, 275 Winchester Avenue, New Haven, Connecticut 06535, USA.

pDHFR-I is available from Dr. K.L.Berkner, ZymoGenetics Inc., 4225 Roosevelt Way NE, Seattle, Washington 98105. (The construction of this plasmid is given in detail in: Berkner, K.L. and Sharp, P.A. (1984) Nucleic Acids Res. 12: 1925-1941). The molecular cloning of the DHFR cDNA present in this plasmid, and its sub-cloning in mammalian expression vectors under the control of adenovirus derived promoters has previously been described in detail (Chang, A.C.Y., et al., Nature 275: 617-624 and Kaufman, R.J. and Sharp, P.A. (1982) Mol. Cell. Biol. 2: 1304-1319) . The backbone plasmid in pDHFR-I is pBR322 (Sutcliffe, J.G. (1979) Cold Spring Harbor Symp. Quant. Biol. 43: 77-90; Sutcliffe, J.G. (1978) Nucleic. Acids Res. 5: 2721-2728).

pUC13 is described in: Vieira, J. and Messing, J.: 1982, Gene 19: 259-268 and available from Pharmacia LKB Biotechnology (catalog # 27-4954-01).

pUC19 is described in: Yanisch-Perron, C. and Messing, J., 1985, Gene 33:103-119 and available from Pharmacia LKB Biotechnology (catalog # 27-4951-01).

Growth media

LB-broth:

Mix 227 g Bacto Tryptone, Difco 0123-01
113.5 g Yeast extract, Difco 0127-01, and
227 g NaCl in a sealable plastic container.

Add 12.5 g mix to 500 ml water in a 1000 ml bottle, shake well and sterilize in an autoclave.

Dulbeccos Modified Eagle Medium is available from e.g. Gibco Ltd. P.O. Box 35, Trident House, Renfrew Road, Paisley PA34EF, Renfrewshire, Scotland. Cat.# 042-250 1M (10 * concentrate).

Antibodies

Anti-α_2M A033 and peroxidase conjugated anti-α_2M PE326 were from DAKOPATTS A/S, Copenhagen, Denmark.

EXAMPLE 1.

CLONING AND SEQUENCE DETERMINATION OF HUMAN α_2M

Preparation of messenger RNA from the human cell line HepG2.

The human hepatoblastoma cell line HepG2 (American Type Culture Collection No. HB 8065, freely available) was used as a source for mRNA preparation. HepG2 cells were grown to a total cell number of 15 * 10^7 in Dulbecco's Modified Eagle medium containing 10% fetal calf serum and antibiotics.

Total RNA was isolated by the guanidinium thiocyanate method (Chirgwin et al., (1979) Biochemistry 18: 5293-5299) and purified by CsCl gradient centrifugation. A total of 3000 µg RNA was obtained. mRNA was isolated by use of an oligo(dT)-cellulose column (Aviv & Leder (1972) Proc. Natl. Acad. Sci. USA 69: 1408-1412). 60 µg of mRNA was obtained after one cycle of affinity chromatography. After ethanol precipitation, this preparation of mRNA was resuspended in 10 mM Tris-HCl pH 7.5, 0.1 mM EDTA-Na_2 at a final concentration of 1 µg/µl and stored at -80°C for subsequent use in the construction of a cDNA library.

Construction of a cDNA library from HepG2 mRNA.

A cDNA library was constructed in the pCDVI-PL/pSP62-K2 vectors (Noma et al., (1986) Nature, 319: 640-646. Available from Dr. Tasuku Honjo,

Faculty of Medicine, Kyoto University, Kyoto 606, Japan) by use of the methods described by Okayama & Berg (Mol. Cell. Biol. 2: 161-170 (1982); Mol. Cell. Biol. 3: 280-289 (1983)).

E. coli K12 (MC1061) (Casadaban & Cohen (1980) J. Mol. Biol. 138: 179-207) was used for transformation. MC1061 were grown in L-broth at 37°C to OD_{660}=0.5. Twenty ml were centrifuged, and the pellet was resuspended in 7 ml of ice-cold sterile 0.1 M $CaCl_2$, incubated on ice for 30 minutes, centrifuged briefly, and finally kept in the cold room overnight.

Ninety-five µl suspension of transformation-competent E. coli MC1061 were added per 10 µl of cDNA preparation. The mixture was incubated on ice for 30 minutes, heat-shocked at 43,5°C for 45 seconds, and finally, after addition of L-broth, incubated at 37°C for 30 minutes.

After resuspension, the cells were plated onto L-broth plates containing ampicillin (50 µg/ml) and grown for 8 hrs at 37°C. A total of 2.9 *10^5 individual colonies could be obtained from this library.

Screening of the HepG2 library for cDNA clones encoding human α_2M.

5 * 10^4 individual colonies were screened by standard colony hybridization technique using nitrocellulose filters (Maniatis et al., (1982) Molecular Cloning - A Laboratory Manual, Cold Spring Harbor, New York).

A 20-mer oligonucleotide mixture

5′ CC(T/C)TTCAT(G/A)TC(T/C)TC(T/C)TG(T/C)TT 3′

where the notation (X/Y) means that either of the nucleic acids X or Y may be used, complementary to the human α_2M mRNA in the region encoding amino acid residues Lys-Gln-Glu-Asp-Met-Lys-Gly (residues number 493 - 499 in Sottrup-Jensen et al., J. Biol. Chem. 259: 8318-8327 (1984) was synthesized (on a DNA synthesizer from Applied Biosystems, USA), labelled with ^{32}P (using T_4 polynucleotide kinase and γ-^{32}P-ATP) to a specific activity of 3 * 10^8 cpm/pmol oligonucleotide. The labelled oligonucleotides were purified by gel chromatography and subsequently used in the screening of the cDNA library.

The hybridization solution contained 6 * SSC, 5 * Denhardt's solution, 0.05% SDS (Maniatis et al., (1982) Molecular Cloning - A Laboratory Manual, Cold Spring Harbor, New York) and 10^6 cpm/ml of labelled oligonucleotide mix.

Hybridization was performed for 3 hrs at 45°C. Then the filters were washed in 6 * SSC, 0.05% SDS at 45°C for 3 * 10 minutes. After autoradiography the filters were washed under the same conditions, but this time at 52°C. A colony that still showed hybridization at this temperature was isolated and the cDNA insert of the corresponding plasmid (designated pα_2M)

from this isolate was sequenced (Tabor & Richardson (1987) Proc. Natl. Acad. Sci. USA 84: 4767-4771). The sequence of the cDNA and the derived encoded amino acid sequence are shown in the appended sequence listings, SEQ ID NO:1:, and SEQ ID NO:2:.

Characterization of $p\alpha_2M$.

$p\alpha_2M$ had a cDNA insert of approximately 4.6 kb. Its sequence is given in Table I above.

The sequence in Table I demonstrates that the entire coding region of α_2M including the signal peptide is found in the insert.

In addition to the coding region, the insert contains sequences derived from the 5′- and 3′ untranslated regions of the α_2M mRNA molecule.

The amino acid sequence of the human α_2M as deduced from the cDNA in $p\alpha_2M$ is in total agreement with the published sequence (Sottrup-Jensen et al., (1984) J. Biol. Chem. 259: 8318-8327). Codon number 1000 (numbered from the initiating methionine codon in the signal peptide) was found to be ATC encoding an isoleucine and not GTC (encoding a valine) as found in an α_2M cDNA synthesized from human liver mRNA (Kan et al., (1985) Proc. Natl. Acad. Sci. USA. 82: 2282-2286). In the α_2M cDNA sequence from the HepG2 library we have further identified ten silent changes as compared to the sequence from the liver library, see the following Table I:

TABLE I

Codon	Liver	HepG2
413 (Asn)	AAC	AAT
495 (Phe)	TTT	TTC
750 (Gly)	GGG	GGT
796 (Leu)	CTT	CTC
835 (Leu)	CTT	CTA
1266 (Ala)	GCC	GCA
1296 (Asn)	AAT	AAC
1326 (Thr)	ACC	ACA
1442 (Leu)	CTC	CTG
1460 (Ile)	ATC	ATT

The position of the oligonucleotide mixture used as a hybridization probe in the colony screenings was from position 1574 to position 1594, and the position of the reactive thiol ester is from position 2939 to 2953 in SEQ ID NO:1.

EXAMPLE 2.

Construction of a mammalian expression vector for α_2M.

pα_2M was digested (fig. 1a) with XbaI and EcoRI, and a 1.2 kb fragment containing the 5′ part of the α_2M cDNA together with the multiple cloning site of pSP62-K2 was isolated on an agarose gel and cloned in an XbaI/EcoRI digested M13mp19 vector to generate M13mp19A. To facilitate further subclonings of the α_2M cDNA, a unique EcoRV site was introduced in the 1.2 kb fragment 10 nucleotides 5′ to the initiating ATG (methionine) codon through site directed mutagenesis (Kunkel et al., (1987) Methods Enzymol. 154: 367-382). In the same mutagenesis experiment, in which the mutagenic oligonucleotide NOR593:

5′(TTCTTCCCCATGGTGGATATCGAAGGAGCTG)3′

was used, the 5 nucleotides 5′ to the methionine codon was changed to CCACCATG; this mutation creates a new NcoI site spanning the ATG codon. A

correct mutant M13mp19B was identified through restriction enzyme digestion and DNA sequencing.

The mutated 5′ end of α_2M cDNA was isolated from M13mp19A replicative form through digestion with HindIII and EcoRI and agarose gel electrophoresis. The isolated DNA fragment was then joined to HindIII/EcoRI digested pα_2M through ligation to generate p1136. In this plasmid the α_2M cDNA is reassembled in its total length, but now with a unique EcoRV site at the 5′ end. p1136 was digested with EcoRV/DraI, and the α_2M fragment was isolated on an agarose gel and cloned in a mammalian expression vector under control of the adenovirus 2 major late promoter (Ad 2 MLP).

The adenovirus-promoter based vector was constructed by K.L.Berkner (ZymoGenetics Inc., Seattle, WA.), and a detailed description of the functional elements in the mammalian expression vector is given in: Powell, J.S. et al., (1986) Proc. Natl. Acad. Sci. USA 83: 6465-6469 and in: Boel et al., (1987) FEBS Lett. 219: 181-188).

The expression vector used for expression of human α_2M was generated from the mammalian expression vector pPP (Boel, E. et al., (1987) FEBS Lett. 219: 181-188), in which human pancreatic polypeptide cDNA was cloned under control of Ad 2 MLP.

pPP was digested (fig. 1b) with BamHI and the resulting staggered ends were repaired with DNA polymerase (Klenow fragment and the four deoxynucleotide triphosphates). The 4.5 kb EcoRV/DraI α_2M cDNA fragment was joined to this vector through ligation, and correct recombinants were characterized through restriction enzyme analysis on isolated miniprep. plasmids.

The α_2M-mRNA transcribed from the resulting 8.76 kb plasmid (designated p1167 (fig. 2)) has the adenovirus 2 late tripartite leader (L1-3) at its 5′ end together with an mRNA splice signal (SS). At the 3′ end of the construct the transcript is terminated with the SV40 late termination - and polyadenylation signal. 5′ to the Ad 2 MLP the construct includes the SV40 enhancer (ENH) and the 0 to 1 (0 - 1) map units from adenovirus 5.

Expression of α_2M in mammalian cells.

For expression of human α_2M in cultured BHK cells (Syrian Hamster Kidney, thymidine kinase mutant line tk$^-$s13, (Waechter and Baserga (1982) Proc. Natl. Acad. Sci. USA 79: 1106-1110); American Type Culture Collection CRL 1632) the expression vector p1167 was co-transfected with pDHFR-I (Berkner, K.L. and Sharp, P.A. (1984) Nucleic Acids Res. 12: 1925-1941. Available from K.L.Berkner, ZymoGenetics Inc. Seattle) into subconfluent cells by the

calcium phosphate mediated transfection procedure (Graham and Van der Eb (1973) Virology 52: 456-467). In the transfection experiment the molar ratio between p1167 and pDHFR-I was 10:1. Cells were grown in Dulbeccos Modified Eagle Medium supplemented with 10% fetal calf serum (FCS).

Forty-eight hours after transfection, cells were trypsinized and diluted into medium containing 400 nM methotrexate (MTX). After 10 to 12 days, individual colonies were cloned out and expanded separately. The expanded cultures were propagated for 24 hours as described above, and producer clones were identified using an enzyme linked immunosorbent assays (ELISA) (Munck Petersen C., et al., (1985) Scand. J. Clin. Lab. Invest. 45: 735-740) against human α_2M secreted to the growth medium.

Description of the α_2M ELISA assay.

The materials used in the ELISA were:

Catching antibody A033 anti-α_2M,

Peroxidase-conjugated anti-α_2M antibody PE326,

1,2-Phenylenediamine, dihydrochloride (OPD)

all from DAKOPATTS A/S, Copenhagen, Denmark.

Urea peroxide, 125 mg, was from Organon Teknika.

96 well ELISA plates were from NUNC, Copenhagen.

Coating buffer:

100 mM carbonate buffer pH 9.6 was made up as follows:

Add 3.18 g Na_2CO_3 and 5.96 g $NaHCO_3$ to 1000 ml water.

Standard and sample buffer:

To 100 ml of 150 mM phosphate buffer pH 7.2 was added:

50 µl Tween 20

2 g Bovine Serum Albumin (Sigma A 7030).

Washing buffer:

10 mM sodium phosphate pH 7.4

145 mM sodium chloride

0.1 % Tween 20.

Citric acid-phosphate buffer, pH 4.9:

The following reagents were added to 1000 ml of water

7.3 g citric acid

23.88 g Na_2HPO_4, 12 H_2O

0.5 ml Tween 20

The buffer was used for a maximum of 14 days, stored at 4°C.

Urea peroxide solution:

125 mg urea peroxide was dissolved in 8.93 ml water.

The solution was kept in the dark at 4°C.

Coating of the plates for assay:

The 96 well plate was coated with 175 µl of the DAKO A033 antibody diluted 1:1000 in the coating buffer. The plate was incubated over night at 4°C. Before use the plate was washed 4 times in washing buffer.

Application of standards and samples:

100 µl standard or sample was added to each well. As a standard purified human α_2M, 2 mg/ml (prepared as described in: Sottrup-Jensen et al., (1983) Ann. N.Y. Acad. Sci. 421: 41-60) was used. The standard curve included the following serial dilutions: 1:4000, 1:8000, 1:16000 etc. down to 1:1024000, corresponding to final concentrations from 500 µg/l down to 1.95 µg/l. All dilutions were done in the Standard and sample buffer. The plate was incubated over night at 4°C and then washed 4 times with wash buffer before the next step.

Addition of conjugated antibody:

100 µl of PE326, which had been diluted 1:6000 in the Standard and sample buffer, was added to each well. The plate was incubated for 2 h at 20°C, and then washed 4 times with wash buffer.

Enzyme activation:

8 mg of OPD was dissolved in 12 ml of Citric acid- phosphate buffer. To this solution 500 µl Urea peroxide solution was added and the mixture was used immediately. 100 µl of the final solution was added to each well, and the plate was incubated in the dark for 6 min. Then 100 µl of 2 M H_2SO_4 was added to each well and the A_{492} was read in an automated ELISA plate reader.

The above described ELISA did not give any background on medium supplemented with 10% FCS, nor did it give any background in BHK cell conditioned medium. Of 24 isolated MTX resistant clones, 16 produced detectable amounts of recombinant α_2M.

Selected cell lines that secreted 12.3 mg/l (K16-6) and 19.1 mg/l (K17-6) in the supernatant (grown in a 6 well NUNC-plate) over a 48 hour period were expanded for large scale production of recombinant human α_2M (rα_2M).

Purification of recombinant human α_2M.

Cell lines K16-6 and K17-6 were each expanded into one ten-double tray (NUNC, Denmark) with a growth surface of 6000 cm^2. At 80% confluency the medium on the cells was changed from containing the 10% fetal calf serum (FCS) down to 2%. After 48 hours of growth in medium with only 2% (FCS), the medium was removed, and the cells were washed twice with serum free medium. Cells were then grown serum free for 4 to 5 days with change of serum free medium every two days. Conditioned medium was pooled and analyzed for rα_2M by ELISA.

The pooled conditioned medium from K16-6 and from K17-6 contained 7.15 mg/l and 21.5 mg/l of rα_2M, respectively.

The rα_2M was purified according to published procedures (Sottrup-Jensen et al., (1983) Ann. N. Y. Acad. Sci. 421: 41-60). Briefly the conditioned medium was loaded onto a 10 ml Zn-Chelate column (Zn^{2+}-iminodiacetic acid Sepharose 4B (Porath, J. et al., (1975) Nature 258: 598-599) equilibrated with 25 mM Tris-HCl pH 8.0, and washed with 100 ml phosphate buffered saline (PBS) pH 7.2 until $A_{280} < 0.036$. A second wash with 20 mM sodium phosphate, 500 mM NaCl pH 6.2 was performed until $A_{280} < 0.033$. The flow rate was 100 ml/hr and 3 ml fractions were collected. rα_2M was eluted with 100 mM EDTA pH 7.0 at a flow rate of 40 ml/hr. During elution 1 ml fractions were collected.

Recovery of rα_2M was 44%. The rα_2M containing fractions were concentrated to 1 ml on an Amicon devise equipped with a PM 10 membrane and then loaded onto a Superose 12 gelfiltration column (25 mM Tris-HCl, 150 mM NaCl pH 8.0). The rα_2M containing fractions were pooled and stored at -20°C until analysis.

EXAMPLE 3.

Characterization of recombinant human rα_2M.

A. Chemical reactions at the thiol ester: thermal fragmentation and methylamine induced cleavage.

A number of different analyses were performed to evaluate the structural and biological characteristics of the human rα_2M as compared to a preparation of human plasma derived α_2M, designated preparation LSJ39.

An important structural feature of α_2M is the presence of the thiol ester. When heated to 95°C for 15 min, the thiol ester will induce a peptide bond cleavage in the backbone of α_2M at the position of the thiol esterified Glx-residue. This results in the fragmentation of the 180 kD α_2M monomer into two polypeptides of 120 kD and 60 kD. Fig. 3 shows an analysis of both the purified rα_2M (from two transformed BHK cell lines) and the purified human plasma derived preparation LSJ39 on a 10-20% SDS polyacrylamide gel. The different preparations, either native human or BHK cell derived recombinant α_2M were all heat treated to induce thermal fragmentation before loading onto the gel. Molecular weight markers (from top to bottom: 180, 120, 92, 60, 43, 26, 14 and 6 kD) were applied to lanes 1 and 8. Samples in lanes 2, 3 and 4 were not reduced before electrophoresis, while samples in lanes 5, 6 and 7 were reduced. Preparation LSJ39 was applied to lanes 2 and 5. rα_2M K16-6 was applied to lanes 3 and 6, and rα_2M K17-6 was applied to lanes 4 and 7.

It was clear from the patterns of protein fragments on the gel, that both human α_2M and the two rα_2M preparations showed a considerable degree of thermal fragmentation. As expected, only the reduced samples displayed this fragmentation. In the nonreduced samples, the molecules migrated as the 360 kD dimer.

In the human plasma derived preparation LSJ39 (lane 5) a fragment migrating slightly faster than the 60 kD fragment could be observed. Lanes 6 and 7 indicated the presence in the recombinant material of a similar faster migrating fragment. It is possible that this fragment represented a slightly underglycosylated variant of the 60 kD fragment.

Methylamine (MA) and other small nitrogen containing nucleophiles will cleave the thiol ester and thereby inactivate the ester (Sottrup-Jensen, L., et al., (1980) FEBS Lett. 121: 275-280; Salvesen, G.S. et al., (1981) Biochem. J. 195: 453-461). After MA induced inactivation of the thiol ester, thermal fragmentation of α_2M can no longer be observed.

Fig. 4 shows a SDS-PAGE run similar to that shown in Fig. 3 (with respect to loaded samples), in which applied α_2M and rα_2M had been pretreated with MA. From this gel it was concluded, that the thiol ester of rα_2M was just as susceptible to cleavage with MA as the thiol ester of native α_2M. Upon reduction MA-treated α_2M and rα_2M migrated as a single 180 kD monomer species.

Lanes 5 of both Fig. 3 and 4 shoved an additional band of approximately 85 kD. When α_2M is cleaved in the bait region by proteinases present in the blood, it generates two fragments, each with a molecular weight of 85 kD. The human α_2M preparation LSJ39 (purified from serum) contained these cleavage products, while they could not be detected on this gel in the two rα_2M preparations. This indicated that the material secreted from the transformed BHK cell lines was largely native uncomplexed α_2M. Any α_2M molecules, that have reacted with proteinases are inactivated and can not form additional complexes with other proteinases. Since the BHK cell does not produce any proteinases that forms complexes with the rα_2M product, this cell is therefore well suited for production of recombinant human α_2M.

B. Reaction with trypsin.

Reaction with trypsin is a standard way of analyzing the proteinase-complex formation ability of α_2M (Sottrup-Jensen, L. (1987) in: "The Plasma Proteins" (Putnam, F.W., ed.) 2nd Ed., 5: 191-291, Academic Press, Orlando, FL; Harpel, P.C. (1973) J. Exp. Med. 138: 508-521; Harpel, P.C., et al., (1979) J. Biol. Chem. 254: 8869-8878; Swenson, R.P. and Howard, J.B. (1979) J. Biol. Chem. 254: 4452-4456). In this reaction trypsin will cleave at its target site(s) in the bait region of α_2M, and the resulting reduced cleavage products (85 kD) will migrate as a double band. Under nonreducing conditions the trypsin-α_2M complexes will migrate as high molecular weight products.

Fig. 5 shows the result of such an analysis (performed as described (Sottrup-Jensen, L. (1987) in: "The Plasma Proteins" (Putnam, F.W., ed.) 2nd Ed., 5: 191-291, Academic Press, Orlando, FL; Harpel, P.C. (1973) J. Exp. Med. 138: 508-521; Harpel, P.C., et al., (1979) J. Biol. Chem. 254: 8869-8878; Swenson, R.P. and Howard, J.B. (1979) J. Biol. Chem. 254: 4452-4456)) on the native human α_2M preparation LSJ39 (lanes 2 and 5) and on rα_2M from cell lines K16-6 (lanes 3 and 6) and K17-6 (lanes 4 and 7). The samples in lanes 2, 3 and 4 were not reduced before electrophoresis, while the samples in lanes 5, 6 and 7 were. Lane 5 shows that almost all of the human native α_2M was cleaved with trypsin, while the two preparations of rα_2M were cleaved with an efficiency of approximately 80% or more. Without reduction of the complexes no low molecular weight products from the reaction between trypsin and the native α_2M or the BHK cell derived rα_2M were seen on the gel. The 85 kD fragments derived from the recombinant material migrated somewhat faster than the human standard; as mentioned above the recombinant material might be slightly underglycosylated.

When α_2M is reacted with methylamine, the thiol ester will be inactivated, and α_2M changes conformation from the "slow" form to the "fast" form (Sottrup-Jensen, L. (1987) in: The Plasma Proteins (Putnam, F.W., ed.) 2nd Ed., 5: 191-291, Academic Press, Orlando, FL; Van Leuven, F., Cassiman, J.-J. and Van Den Berghe, H. (1981) J. Biol. Chem. 256: 9016-9022). In this conformation it can no longer react rapidly with or form complexes with proteinases such as e.g. trypsin.

Fig. 6 shows the results of a set of experiments that were run in parallel to the experiments described above and shown in Fig. 5. However, before reaction with trypsin the native human α_2M and the rα_2M used in this experiment had been treated with methylamine (Sottrup-Jensen, L., et al., (1980) FEBS Lett. 121: 275-280). Under these conditions both the native α_2M and the rα_2M show a marked decrease in reactivity towards trypsin (80% or more of the α_2M and rα_2M monomers were migrating as a 180 kD polypeptide). This indicates that trypsin does not rapidly cleave at the bait region in methylamine treated human α_2M or in BHK cell derived rα_2M.

In these types of experiments BHK cell derived rα_2M has shown characteristics similar to those of native human α_2M.

C. Trypsin and methylamine induced conformational change in α_2M.

As mentioned above the α_2M molecule will undergo a conformational change both through complex formation with proteinases and through methylamine induced cleavage of the thiol ester. The change in structure results in an altered mobility on rate gels (Sottrup-Jensen, L. (1987) in: The Plasma Proteins (Putnam, F.W., ed.) 2nd Ed., 5: 191-291, Academic Press, Orlando, FL; Van Leuven, F., Cassiman, J.-J. and Van Den Berghe, H. (1981) J. Biol. Chem. 256: 9016-9022); unreacted α_2M will migrate as a "slow" form, while reacted α_2M will migrate as a "fast" form.

Fig. 7 and Fig. 8 show these conformational changes, as they appear after reaction with trypsin and methylamine, respectively (analyzed on 5-10% rate gels).

Lanes 1 on both gels contain purified human pregnancy zone protein (PZP) (Sand, O. et al., (1985) J. Biol. Chem. 260: 15723-15735), which is known to appear in both a dimeric (D) and a tetrameric (T) configuration.

Lanes 2 on both gels contain unreacted human α_2M preparation LSJ39. Lanes 3 on both gels show the fast migrating form, resulting from reaction with trypsin and methylamine, respectively. Lanes 4 on both gels show the unreacted rα_2M preparation K16-6, and lanes 5 show the corresponding

fast forms. Lanes 6 on both gels show the unreacted rα_2M preparation K17-6, and lanes 7 show the corresponding fast forms.

It can be concluded that both complex formation between rα_2M and trypsin and reaction of rα_2M with methylamine result in the appearance of fast migrating structures. These structures appear (as analyzed on rate gels) to be very similar to the structures obtained when human α_2M was allowed to react with trypsin and methylamine. It is also evident from these figures that the rα_2M proteins showed a migration, which, when compared to the migration of dimeric and tetrameric PZP on the gels, is in agreement with the finding that these molecules are produced and secreted from the BHK cells in the active tetrameric conformation.

D. Chromatography of α_2M on a Superose 6 column.

A Superose 6 column can partially resolve α_2M molecules in the dimeric configuration from molecules in the tetrameric configuration (Sottrup-Jensen, L. unpublished). Human standard α_2M and rα_2M was analyzed on a 24 ml Superose 6 column (buffer: 25 mM Tris-HCl, 125 mM NaCl pH 8.0; flow rate: 1 ml/min; fraction size: 1 ml). Fig. 9 shows the diagrams obtained from the chromatography of purified human standard α_2M and rα_2M from the K17-6 and the K16-6 BHK cell lines. Tetrameric α_2M (Sottrup-Jensen, unpublished observation) will elute in fraction 12 on this type of column. It is evident from the chromatograms that both of the rα_2M preparations eluted in fraction 12, as did the human standard α_2M. On this type of column, dimeric α_2M molecules will elute in fraction 14 and 15 (Sottrup-Jensen, unpublished observation). This type of analysis supported the results obtained from the rate gels (Figs. 7 and 8), that rα_2M was secreted from BHK cells in a tetrameric configuration.

E. Trypsin protection analysis.

When trypsin is trapped inside the α_2M molecule, it retains its catalytic capacity towards low molecular weight substrates such as S-2222 (N-benzoyl-L-Ile-L-Glu-Gly-L-Arg-p-nitroanilide). If trypsin is efficiently complexed with α_2M, it will be protected against high molecular weight inhibitors such as Soybean Trypsin Inhibitor (STI) (Sottrup-Jensen, L. (1987) in: The Plasma Proteins (Putnam, F.W., ed.) 2nd Ed., 5: 191-291, Academic Press, Orlando, FL; Ganrot, P.O. (1966) Clin. Chim. Acta. 14: 493-501; Sottrup-Jensen, L. et al., (1981) FEBS Lett. 128: 127-132).

K16-6 and K17-6 derived rα_2M was compared with human plasma α_2M in such a protection assay. 100 μl α_2M (in 25 mM Tris-HCl, 125 mM NaCl, pH

8.0) was mixed with 30 µl trypsin (0.5 mg/ml in 20 mM sodium acetate pH 5.0). After incubating for 2 min. 30 µl 1 mg/ml STI (in PBS) was added. 10 µl aliquots were removed after 2 and 4 min. and each mixed with 750 µl 0.12 mM S-2222 (dissolved 0.1 M sodiumphosphate pH 8.0, 5% dimethylsulfoxide).

The change in absorbance at 405 nm was recorded for 2 min. The results of the assay are given in the following Table II:

TABLE II

Prep. of α_2M.	α_2M in cuvette.		Activity.
	A_{405}/min	µg	A_{405}/min/µg
Human LSJ39	0.140	5.00	0.028
K16-6	0.111	4.62	0.024
K17-6	0.119	4.87	0.024

From these results it can be concluded that rα_2M had essentially the same protection capacity for trypsin against STI as compared with the protection capacity of human plasma α_2M.

If α_2M is treated with methylamine before the protection assay, the protection capacity drops dramatically. In a similar assay as that described above, methylamine treated human plasma α_2M only retained 17% of its protection capacity, while K16-6 and K17-6 rα_2M retained 16% and 14% respectively. It can be concluded that rα_2M protected trypsin against STI with almost the same efficiency as did human plasma α_2M.

E. Amino terminal amino acid sequencing of rα_2M.

Theoretically, the α_2M characterized in the present investigation could only be either bovine (contaminant from serum), from hamster (endogenous product from the BHK cell) or derived from expression of the transfected plasmid p1167. The ELISA assay used never recognized any α_2M in BHK cell conditioned medium, whether with or without added fetal calf serum. To make sure that the investigated α_2M was human α_2M, and to characterize the amino terminal processing of the recombinant product, amino terminal amino acid sequence determination was carried on out K16-6 and K17-6 rα_2M as described (Sottrup-Jensen, L. et al., (1984) J. Biol. Chem. 259: 8293-8303). The Edman degradation was repeated for 12 cycles, and the identity of the detected amino acid derivative in each cycle, was in total agreement with the

amino terminal sequence of human α_2M: Ser-Val-Ser-Gly-Lys-Pro-Gln-Tyr-Met-Val-Leu-Val-, whereas bovine α_2M has the following amino terminal sequence: Ala-Val-Asp-Gly-Lys-Pro-Gln-Tyr-Met-Val-Leu-Val- (unpublished, Dr. Torsten Kristensen, Department of Molecular Biology, University of Aarhus, Denmark.)

EXAMPLE 4.

Construction and expression of a bait region mutant of human α_2M.

In the present example it is demonstrated that the bait region of human α_2M can be substituted by the bait region of human pregnancy zone protein (PZP) (Sottrup Jensen, L., Folkersen, J., Kristensen, T. and Tack, B.F. Partial primary structure of human pregnancy zone protein: extensive sequence homology with human alpha 2-macroglobulin. Proc. Natl. Acad. Sci. U.S.A. 81, 7353-7357, 1984; Sand, O., Folkersen, J., Westergaard, J.G. and Sottrup Jensen, L. Characterization of human pregnancy zone protein. Comparison with human alpha 2-macroglobulin. J.Biol.Chem. 260, 15723-15735, 1985). The resulting α_2M bait region mutant exhibited a proteinase inhibitor profile similar to that of human pregnancy zone protein.

To facilitate substitution of DNA fragments encoding the bait region of human α_2M cDNA, target sites for the restriction enzymes PstI and SacII were introduced at the 5′ and at the 3′ end of the cDNA region encoding the bait region.

The human α_2M expression plasmid p1167 was digested with BamHI and ClaI, and a 2660 bp fragment, which carried the central part of the human α_2M cDNA, was subcloned in the BamHI and ClaI digested vector pSX191.

This vector, which had previously been constructed, is a derivative of pUC19. It was constructed as described: pUC19 was digested with EcoRI and HindIII, and a synthetic linker with the following sequence

```
     KpnI  PstI  EcoRI Hind3 ClaI   SphI  BamHI
AATTGGTACCCTGCAGGAATTCAAGCTTATCGATGGCATGCGGATCC     - NOR781
    CCATGGGACGTCCTTAAGTTCGAATAGCTACCGTACGCCTAGGTCGA - NOR782
```

was cloned in the digested pUC19 vector. The linker, which was an annealing product from the two synthetic oligonucleotides NOR781 and NOR782, has cohesive ends that will ligate to the EcoRI and the HindIII sites of pUC19 in such a way that these ligation sites are not regenerated in the pSX191 vector. Thus pSX191 carried sites for KpnI, PstI, EcoRI, HindIII, ClaI, SphI and BamHI.

The resulting plasmid pSX191α_2M was digested with BamHI and HindIII, and a purified 2.6 kb BamHI/HindIII α_2M fragment was cloned in

M13mp18 to generate M13mp18α_2M for mutagenesis by described methods. A synthetic oligonucleotide NOR973, with the following sequence:

5'(TTCATACTGCTGCAGCTGTGGACAC)3'

was used to introduce a PstI site at position 2102 (SEQ ID NO:1) in the cDNA sequence, and a oligonucleotide (NOR974) with the following sequence:

5'(AGCCACCCCCGCGGAGTTTACCAC)3'

was used to introduce a SacII site at position 2271 (SEQ ID NO:1) in the cDNA sequence. These sites were chosen because they did not introduce alterations in the encoded amino acid sequence, and they were within a convenient distance of the bait region in human α_2M cDNA. Both primers were used in the same mutagenesis experiment (Kunkel, T.A., Roberts, J.D. and Zakour, R.A. Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection. Methods in Enzymol. 154, 367-382, 1987); dsDNA was isolated from mutated M13mp18α_2M plaques, and the DNA was digested with the restriction enzymes PstI and SacII. Correctly mutated recombinants, which had an insert of 160 bp, were further analyzed by DNA sequencing (Tabor, S. and Richardson, C.C. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 84, 4767-4771, 1987). A 2.6 kb BamHI/HindIII fragment from a correct α_2M cDNA mutant (M13mp18α_2M#212.1) was subcloned in a BamHI/HindIII digested pUC13 vector, and a correct subclone p1308 was isolated and characterized with BamHI/HindIII and PstI/SacII double digestions and DNA electrophoresis.

The PstI/SacII fragment in p1308 can be excised and replaced with a different DNA fragment, which encodes bait region variants. The resulting new variants (bait region mutants or analogs) of α_2M cDNA can be isolated as BamHI/ClaI fragments and subcloned back into BamHI/ClaI digested expression vector p1167.

In the present example DNA encoding the amino acids of the bait region for human PZP (Sottrup-Jensen et al. 1989, supra) was obtained from ligation, annealing and cloning of 8 synthetic oligonucleotides.

The DNA sequence of the synthetic fragment and the encoded amino acids as inserted into the α_2M clone are given in SEQ ID NO:3, and comprises positions 2107 to 2305 and the corresponding amino acids. A PstI site was introduced at the 5' end in the synthetic fragment, and SacII and BamHI sites were introduced at the 3' end.

This synthetic 0.2 kb DNA fragment was cloned in a PstI/BamHI digested M13mp18 vector for DNA sequencing. DNA from a clone containing the correct sequence was digested with PstI and SacII, and the purified 0.2 kb fragment was cloned in a PstI/SacII digested and gel purified p1308 vector.

A correct recombinant, p267PZP, was characterized with restriction enzyme digestions, and from this plasmid, bait region mutated (α_2M → PZP) cDNA was isolated as a 2.7 kb BamHI/ClaI fragment and subcloned in a BamHI/ClaI digested α_2M expression vector p1167. The resulting plasmid, designated p1365, was grown as a large scale plasmid preparation, purified by CsCl centrifugation, and cotransfected with pDHFR-I into BHK cells.

Through this procedure the nucleotides 2102 to 2275 in SEQ ID NO:1 was removed and replaced with nucleotides 2102 to 2305 in SEQ ID NO:3.

The procedures for transfection, selection of bait region mutated α_2M (designated rα_2M-PZP) recombinants (with an α_2M specific ELISA), large scale production and purification of mutated α_2M were as described elsewhere (EXAMPLE 2) in this application.

Characterization of the proteinase inhibitor specificity of a bait region mutant of human α_2M.

The purified recombinant α_2M mutant, rα_2M-PZP, was characterized with respect to its inhibitor specificity profile against various proteinases by the use of previously described methods (Sand et al.1985). For comparison human plasma derived α_2M and PZP were treated with the same set of proteinases in parallel reactions. The proteinases used were chymotrypsin, elastase, trypsin and Staphylococcus aureus Glu-specific proteinase. It has been reported (Sand et al.1985) that chymotrypsin and elastase show a rapid reaction with both PZP and α_2M, while the reaction between the two proteinase inhibitors and trypsin and Staphylococcus aureus Glu-specific proteinase is quite dissimilar for PZP and α_2M: both proteinases react rapidly with α_2M, while the reaction with PZP is slow (Sand et al.1985). The reason for this difference in reaction rate with the different proteinases is believed to be due to the fact that the bait region in PZP contains strong specificity determinant for chymotrypsin and elastase, but none for trypsin and Staphylococcus aureus Glu-specific proteinase.

The results of the analysis is presented in figures 10 to 13.

Figure 10 illustrates the gel electrophoresis (10 - 20 % reducing SDS-PAGE) of the reaction products from chymotrypsin treated human α_2M, human PZP and rα_2M-PZP. Molecular weight markers (from top to bottom: 180, 120, 92, 60, 43, 26, 14 and 6 kD) were applied to lanes 1 and 8. All samples were reduced. Lanes 2, 3 and 4 show the cleavage products obtained from reaction of chymotrypsin with human plasma derived PZP, rα_2M-PZP and human plasma derived α_2M, respectively. The ratio of proteinase to inhibitor was 1:1. Lanes 5, 6 and 7 show cleavage products from similar reactions at a ratio of 2:1

between proteinase and the three tested inhibitors. In all 6 lanes cleavage products (85 kD) could be identified. This indicated that rα_2M-PZP reacted with chymotrypsin with similar characteristics as did human plasma derived α_2M and PZP.

Figure 11 illustrates the gel electrophoresis (10 - 20 % reducing SDS-PAGE) of the reaction products from elastase treated human α_2M, human PZP and rα_2M-PZP. Molecular weight markers were the same as applied on the gel in Fig. 2. All samples were reduced. Lanes 2, 3 and 4 show the cleavage products obtained from reaction of elastase with human plasma derived PZP, rα_2M-PZP and human plasma derived α_2M, respectively. The ratio of proteinase to inhibitor was 1:1. Lanes 5, 6 and 7 show cleavage products from similar reactions at a ratio of 2:1 between proteinase and the three tested inhibitors. In all 6 lanes cleavage products (85 kD) could be identified. This indicated that rα_2M-PZP reacted with elastase with similar characteristics as did human plasma derived α_2M and PZP.

Figure 12 illustrates the gel electrophoresis (10 - 20 % reducing SDS-PAGE) of the reaction products from trypsin treated human α_2M, human PZP and rα_2M-PZP. Molecular weight markers were the same as applied on the gel in Fig. 2. All samples were reduced. Lanes 2, 3 and 4 show the cleavage products obtained from reaction of trypsin with human plasma derived PZP, human plasma derived α_2M and rα_2M-PZP, respectively. The ratio of proteinase to inhibitor was 1:1. Lanes 5, 6 and 7 show cleavage products from similar reactions at a ratio of 2:1 between proteinase and the three tested inhibitors. In lanes 3 and 6 cleavage products (85 kD) could be identified from the reaction between trypsin and α_2M. In lanes 2, 4, 5 and 7 no cleavage products were observed from the reaction of trypsin with PZP and rα_2M-PZP. This result demonstrated that rα_2M-PZP reacted poorly with trypsin as did human plasma derived PZP, while α_2M was cleaved in the reaction with trypsin.

Figure 13 illustrates the gel electrophoresis (10 - 20 % reducing SDS-PAGE) of the reaction products from Staphylococcus aureus Glu-specific protease treated human α_2M, human PZP and rα_2M-PZP. Molecular weight markers were the same as applied on the gel in Fig. 2. All samples were reduced. Lanes 2, 3 and 4 show the cleavage products obtained from reaction of Staphylococcus aureus Glu-specific protease with human plasma derived PZP, rα_2M-PZP and human plasma derived α_2M, respectively. The ratio of proteinase to inhibitor was 1:1. Lanes 5, 6 and 7 show cleavage products from similar reactions at a ratio of 2:1 between proteinase and the three tested inhibitors. In lanes 4 and 7 cleavage products (85 kD) could be identified from the reaction between Staphylococcus aureus Glu-specific protease and

α_2M. In lanes 2, 3, 5 and 6 much less cleavage product could be identified from the reaction of this proteinase with PZP and rα_2M-PZP. This result demonstrated that rα_2M-PZP reacted poorly with the Staphylococcus aureus proteinase as did human plasma derived PZP, while α_2M was cleaved in the reaction with this proteinase.

It can be concluded that rα_2M-PZP showed the same pattern of reaction with four proteinases as did human plasma derived PZP. This pattern of reaction was different from the corresponding pattern obtained from reaction with α_2M. Thus rα_2M-PZP has been demonstrated to have a proteinase inhibitor profile similar to native PZP and dissimilar to α_2M. Thus it has been demonstrated that the proteinase inhibitor profile of α_2M can be modulated by substitution of DNA fragments encoding the bait region.

The substitution as described in this invention did not destroy the activity of the proteinase inhibitor, and it is therefore demonstrated that functional macroglobulin hybrids can be constructed by substitutions (mutations) in the bait region. The finding will lead to the design of α_2M-derivatives with new desired proteinase specificities. No doubt, these results could be extended to other macroglobulin based hybrids, in which the bait region can be modified at will to obtain new inhibitor specificities.

Aggressive activity of proteinases is often a problem in relation to various diseases (e.g. the activity of elastase and cathepsin G in severe inflammation leads to tissue and organ destruction and failure). Inhibitors of such proteinases will be useful in drug design. In situations where the target site for the proteinase is known, but no inhibitor can be identified, α_2M can be engineered (mutated in the bait region) to obtain the desired specificity. In a situation where the target specificity of the proteinase in question is unknown, saturation mutagenesis or random synthesis of the bait region will lead to an indefinite number of target sequences that can be introduced and expressed in hybrid macroglobulins. These hybrids can be screened for proteinase inhibition, and the target sequence(s) can be identified. The resulting α_2M analog can be produced and purified as described elsewhere in this invention. Upon injection into the circulation such α_2M analogs will inhibit and clear from the blood any proteinase of the given specificity.

Introduction of protein analogs or mutants in the human body always raises the possibility for antigenicity. The generation of a panel of 45 mouse monoclonal antibodies against human α_2M has been described (Van Leuven et al.1988; Delain et al.1988). None of these antibodies were directed against the bait region. This indicates that the bait region is not highly

antigenic and that mutants in this region of the molecule can be generated and used for therapeutical uses without risk for antibody development.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Novo Nordisk A/S

(ii) TITLE OF INVENTION: Expression of Plasma Glycoproteins

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Novo Nordisk A/S, Patent Department
(B) STREET: Novo Alle
(C) CITY: Bagsvaerd
(E) COUNTRY: DENMARK
(F) ZIP: DK-2880

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: DK 4235/89, DK 4236/89, DK 4237/89
(B) FILING DATE: 29-AUG-1989

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4569 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: N

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(F) TISSUE TYPE: Hepatic
(G) CELL TYPE: Hepatoblastoma
(H) CELL LINE: HepG2

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 29..4450
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GTCTCCTCCA GCTCCTTCTT TCTGCAAC ATG GGG AAG AAC AAA CTC CTT CAT          52
                               Met Gly Lys Asn Lys Leu Leu His
                                 1               5

CCA AGT CTG GTT CTT CTC CTC TTG GTC CTC CTG CCC ACA GAC GCC TCA         100
Pro Ser Leu Val Leu Leu Leu Leu Val Leu Leu Pro Thr Asp Ala Ser
     10                  15                  20
```

```
GTC TCT GGA AAA CCG CAG TAT ATG GTT CTG GTC CCC TCC CTG CTC CAC          148
Val Ser Gly Lys Pro Gln Tyr Met Val Leu Val Pro Ser Leu Leu His
 25                  30                  35                  40

ACT GAG ACC ACT GAG AAG GGC TGT GTC CTT CTG AGC TAC CTG AAT GAG          196
Thr Glu Thr Thr Glu Lys Gly Cys Val Leu Leu Ser Tyr Leu Asn Glu
                 45                  50                  55

ACA GTG ACT GTA AGT GCT TCC TTG GAG TCT GTC AGG GGA AAC AGG AGC          244
Thr Val Thr Val Ser Ala Ser Leu Glu Ser Val Arg Gly Asn Arg Ser
             60                  65                  70

CTC TTC ACT GAC CTG GAG GCG GAG AAT GAC GTA CTC CAC TGT GTC GCC          292
Leu Phe Thr Asp Leu Glu Ala Glu Asn Asp Val Leu His Cys Val Ala
         75                  80                  85

TTC GCT GTC CCA AAG TCT TCA TCC AAT GAG GAG GTA ATG TTC CTC ACT          340
Phe Ala Val Pro Lys Ser Ser Ser Asn Glu Glu Val Met Phe Leu Thr
     90                  95                 100

GTC CAA GTG AAA GGA CCA ACC CAA GAA TTT AAG AAG CGG ACC ACA GTG          388
Val Gln Val Lys Gly Pro Thr Gln Glu Phe Lys Lys Arg Thr Thr Val
105                 110                 115                 120

ATG GTT AAG AAC GAG GAC AGT CTG GTC TTT GTC CAG ACA GAC AAA TCA          436
Met Val Lys Asn Glu Asp Ser Leu Val Phe Val Gln Thr Asp Lys Ser
                125                 130                 135

ATC TAC AAA CCA GGG CAG ACA GTG AAA TTT CGT GTT GTC TCC ATG GAT          484
Ile Tyr Lys Pro Gly Gln Thr Val Lys Phe Arg Val Val Ser Met Asp
            140                 145                 150

GAA AAC TTT CAC CCC CTG AAT GAG TTG ATT CCA CTA GTA TAC ATT CAG          532
Glu Asn Phe His Pro Leu Asn Glu Leu Ile Pro Leu Val Tyr Ile Gln
        155                 160                 165

GAT CCC AAA GGA AAT CGC ATC GCA CAA TGG CAG AGT TTC CAG TTA GAG          580
Asp Pro Lys Gly Asn Arg Ile Ala Gln Trp Gln Ser Phe Gln Leu Glu
    170                 175                 180

GGT GGC CTC AAG CAA TTT TCT TTT CCC CTC TCA TCA GAG CCC TTC CAG          628
Gly Gly Leu Lys Gln Phe Ser Phe Pro Leu Ser Ser Glu Pro Phe Gln
185                 190                 195                 200

GGC TCC TAC AAG GTG GTG GTA CAG AAG AAA TCA GGT GGA AGG ACA GAG          676
Gly Ser Tyr Lys Val Val Val Gln Lys Lys Ser Gly Gly Arg Thr Glu
                205                 210                 215

CAC CCT TTC ACC GTG GAG GAA TTT GTT CTT CCC AAG TTT GAA GTA CAA          724
His Pro Phe Thr Val Glu Glu Phe Val Leu Pro Lys Phe Glu Val Gln
            220                 225                 230

GTA ACA GTG CCA AAG ATA ATC ACC ATC TTG GAA GAA GAG ATG AAT GTA          772
Val Thr Val Pro Lys Ile Ile Thr Ile Leu Glu Glu Glu Met Asn Val
        235                 240                 245
```

```
TCA GTG TGT GGC CTA TAC ACA TAT GGG AAG CCT GTC CCT GGA CAT GTG          820
Ser Val Cys Gly Leu Tyr Thr Tyr Gly Lys Pro Val Pro Gly His Val
    250                 255                 260

ACT GTG AGC ATT TGC AGA AAG TAT AGT GAC GCT TCC GAC TGC CAC GGT          868
Thr Val Ser Ile Cys Arg Lys Tyr Ser Asp Ala Ser Asp Cys His Gly
265                 270                 275                 280

GAA GAT TCA CAG GCT TTC TGT GAG AAA TTC AGT GGA CAG CTA AAC AGC          916
Glu Asp Ser Gln Ala Phe Cys Glu Lys Phe Ser Gly Gln Leu Asn Ser
                285                 290                 295

CAT GGC TGC TTC TAT CAG CAA GTA AAA ACC AAG GTC TTC CAG CTG AAG          964
His Gly Cys Phe Tyr Gln Gln Val Lys Thr Lys Val Phe Gln Leu Lys
            300                 305                 310

AGG AAG GAG TAT GAA ATG AAA CTT CAC ACT GAG GCC CAG ATC CAA GAA         1012
Arg Lys Glu Tyr Glu Met Lys Leu His Thr Glu Ala Gln Ile Gln Glu
        315                 320                 325

GAA GGA ACA GTG GTG GAA TTG ACT GGA AGG CAG TCC AGT GAA ATC ACA         1060
Glu Gly Thr Val Val Glu Leu Thr Gly Arg Gln Ser Ser Glu Ile Thr
    330                 335                 340

AGA ACC ATA ACC AAA CTC TCA TTT GTG AAA GTG GAC TCA CAC TTT CGA         1108
Arg Thr Ile Thr Lys Leu Ser Phe Val Lys Val Asp Ser His Phe Arg
345                 350                 355                 360

CAG GGA ATT CCC TTC TTT GGG CAG GTG CGC CTA GTA GAT GGG AAA GGC         1156
Gln Gly Ile Pro Phe Phe Gly Gln Val Arg Leu Val Asp Gly Lys Gly
                365                 370                 375

GTC CCT ATA CCA AAT AAA GTC ATA TTC ATC AGA GGA AAT GAA GCA AAC         1204
Val Pro Ile Pro Asn Lys Val Ile Phe Ile Arg Gly Asn Glu Ala Asn
            380                 385                 390

TAT TAC TCC AAT GCT ACC ACG GAT GAG CAT GGC CTT GTA CAG TTC TCT         1252
Tyr Tyr Ser Asn Ala Thr Thr Asp Glu His Gly Leu Val Gln Phe Ser
        395                 400                 405

ATC AAC ACC ACC AAT GTT ATG GGT ACC TCT CTT ACT GTT AGG GTC AAT         1300
Ile Asn Thr Thr Asn Val Met Gly Thr Ser Leu Thr Val Arg Val Asn
    410                 415                 420

TAC AAG GAT CGT AGT CCC TGT TAC GGC TAC CAG TGG GTG TCA GAA GAA         1348
Tyr Lys Asp Arg Ser Pro Cys Tyr Gly Tyr Gln Trp Val Ser Glu Glu
425                 430                 435                 440

CAC GAA GAG GCA CAT CAC ACT GCT TAT CTT GTG TTC TCC CCA AGC AAG         1396
His Glu Glu Ala His His Thr Ala Tyr Leu Val Phe Ser Pro Ser Lys
                445                 450                 455

AGC TTT GTC CAC CTT GAG CCC ATG TCT CAT GAA CTA CCC TGT GGC CAT         1444
Ser Phe Val His Leu Glu Pro Met Ser His Glu Leu Pro Cys Gly His
            460                 465                 470
```

```
ACT CAG ACA GTC CAG GCA CAT TAT ATT CTG AAT GGA GGC ACC CTG CTG          1492
Thr Gln Thr Val Gln Ala His Tyr Ile Leu Asn Gly Gly Thr Leu Leu
        475                 480                 485

GGG CTG AAG AAG CTC TCC TTC TAT TAT CTG ATA ATG GCA AAG GGA GGC          1540
Gly Leu Lys Lys Leu Ser Phe Tyr Tyr Leu Ile Met Ala Lys Gly Gly
    490                 495                 500

ATT GTC CGA ACT GGG ACT CAT GGA CTG CTT GTG AAG CAG GAA GAC ATG          1588
Ile Val Arg Thr Gly Thr His Gly Leu Leu Val Lys Gln Glu Asp Met
505                 510                 515                 520

AAG GGC CAT TTT TCC ATC TCA ATC CCT GTG AAG TCA GAC ATT GCT CCT          1636
Lys Gly His Phe Ser Ile Ser Ile Pro Val Lys Ser Asp Ile Ala Pro
                525                 530                 535

GTC GCT CGG TTG CTC ATC TAT GCT GTT TTA CCT ACC GGG GAC GTG ATT          1684
Val Ala Arg Leu Leu Ile Tyr Ala Val Leu Pro Thr Gly Asp Val Ile
            540                 545                 550

GGG GAT TCT GCA AAA TAT GAT GTT GAA AAT TGT CTG GCC AAC AAG GTG          1732
Gly Asp Ser Ala Lys Tyr Asp Val Glu Asn Cys Leu Ala Asn Lys Val
        555                 560                 565

GAT TTG AGC TTC AGC CCA TCA CAA AGT CTC CCA GCC TCA CAC GCC CAC          1780
Asp Leu Ser Phe Ser Pro Ser Gln Ser Leu Pro Ala Ser His Ala His
    570                 575                 580

CTG CGA GTC ACA GCG GCT CCT CAG TCC GTC TGC GCC CTC CGT GCT GTG          1828
Leu Arg Val Thr Ala Ala Pro Gln Ser Val Cys Ala Leu Arg Ala Val
585                 590                 595                 600

GAC CAA AGC GTG CTG CTC ATG AAG CCT GAT GCT GAG CTC TCG GCG TCC          1876
Asp Gln Ser Val Leu Leu Met Lys Pro Asp Ala Glu Leu Ser Ala Ser
                605                 610                 615

TCG GTT TAC AAC CTG CTA CCA GAA AAG GAC CTC ACT GGC TTC CCT GGG          1924
Ser Val Tyr Asn Leu Leu Pro Glu Lys Asp Leu Thr Gly Phe Pro Gly
            620                 625                 630

CCT TTG AAT GAC CAG GAC GAT GAA GAC TGC ATC AAT CGT CAT AAT GTC          1972
Pro Leu Asn Asp Gln Asp Asp Glu Asp Cys Ile Asn Arg His Asn Val
        635                 640                 645

TAT ATT AAT GGA ATC ACA TAT ACT CCA GTA TCA AGT ACA AAT GAA AAG          2020
Tyr Ile Asn Gly Ile Thr Tyr Thr Pro Val Ser Ser Thr Asn Glu Lys
    650                 655                 660

GAT ATG TAC AGC TTC CTA GAG GAC ATG GGC TTA AAG GCA TTC ACC AAC          2068
Asp Met Tyr Ser Phe Leu Glu Asp Met Gly Leu Lys Ala Phe Thr Asn
665                 670                 675                 680

TCA AAG ATT CGT AAA CCC AAA ATG TGT CCA CAG CTT CAA CAG TAT GAA          2116
Ser Lys Ile Arg Lys Pro Lys Met Cys Pro Gln Leu Gln Gln Tyr Glu
                685                 690                 695
```

```
ATG CAT GGA CCT GAA GGT CTA CGT GTA GGT TTT TAT GAG TCA GAT GTA      2164
Met His Gly Pro Glu Gly Leu Arg Val Gly Phe Tyr Glu Ser Asp Val
            700                 705                 710

ATG GGA AGA GGC CAT GCA CGC CTG GTG CAT GTT GAA GAG CCT CAC ACG      2212
Met Gly Arg Gly His Ala Arg Leu Val His Val Glu Glu Pro His Thr
        715                 720                 725

GAG ACC GTA CGA AAG TAC TTC CCT GAG ACA TGG ATC TGG GAT TTG GTG      2260
Glu Thr Val Arg Lys Tyr Phe Pro Glu Thr Trp Ile Trp Asp Leu Val
    730                 735                 740

GTG GTA AAC TCA GCA GGT GTG GCT GAG GTA GGA GTA ACA GTC CCT GAC      2308
Val Val Asn Ser Ala Gly Val Ala Glu Val Gly Val Thr Val Pro Asp
745                 750                 755                 760

ACC ATC ACC GAG TGG AAG GCA GGG GCC TTC TGC CTG TCT GAA GAT GCT      2356
Thr Ile Thr Glu Trp Lys Ala Gly Ala Phe Cys Leu Ser Glu Asp Ala
                765                 770                 775

GGA CTT GGT ATC TCT TCC ACT GCC TCT CTC CGA GCC TTC CAG CCC TTC      2404
Gly Leu Gly Ile Ser Ser Thr Ala Ser Leu Arg Ala Phe Gln Pro Phe
            780                 785                 790

TTT GTG GAG CTT ACA ATG CCT TAC TCT GTG ATT CGT GGA GAG GCC TTC      2452
Phe Val Glu Leu Thr Met Pro Tyr Ser Val Ile Arg Gly Glu Ala Phe
        795                 800                 805

ACA CTC AAG GCC ACG GTC CTA AAC TAC CTT CCC AAA TGC ATC CGG GTC      2500
Thr Leu Lys Ala Thr Val Leu Asn Tyr Leu Pro Lys Cys Ile Arg Val
    810                 815                 820

AGT GTG CAG CTG GAA GCC TCT CCC GCC TTC CTA GCT GTC CCA GTG GAG      2548
Ser Val Gln Leu Glu Ala Ser Pro Ala Phe Leu Ala Val Pro Val Glu
825                 830                 835                 840

AAG GAA CAA GCG CCT CAC TGC ATC TGT GCA AAC GGG CGG CAA ACT GTG      2596
Lys Glu Gln Ala Pro His Cys Ile Cys Ala Asn Gly Arg Gln Thr Val
                845                 850                 855

TCC TGG GCA GTA ACC CCA AAG TCA TTA GGA AAT GTG AAT TTC ACT GTG      2644
Ser Trp Ala Val Thr Pro Lys Ser Leu Gly Asn Val Asn Phe Thr Val
            860                 865                 870

AGC GCA GAG GCA CTA GAG TCT CAA GAG CTG TGT GGG ACT GAG GTG CCT      2692
Ser Ala Glu Ala Leu Glu Ser Gln Glu Leu Cys Gly Thr Glu Val Pro
        875                 880                 885

TCA GTT CCT GAA CAC GGA AGG AAA GAC ACA GTC ATC AAG CCT CTG TTG      2740
Ser Val Pro Glu His Gly Arg Lys Asp Thr Val Ile Lys Pro Leu Leu
    890                 895                 900

GTT GAA CCT GAA GGA CTA GAG AAG GAA ACA ACA TTC AAC TCC CTA CTT      2788
Val Glu Pro Glu Gly Leu Glu Lys Glu Thr Thr Phe Asn Ser Leu Leu
905                 910                 915                 920
```

```
TGT CCA TCA GGT GGT GAG GTT TCT GAA GAA TTA TCC CTG AAA CTG CCA      2836
Cys Pro Ser Gly Gly Glu Val Ser Glu Glu Leu Ser Leu Lys Leu Pro
                925                 930                 935

CCA AAT GTG GTA GAA GAA TCT GCC CGA GCT TCT GTC TCA GTT TTG GGA      2884
Pro Asn Val Val Glu Glu Ser Ala Arg Ala Ser Val Ser Val Leu Gly
            940                 945                 950

GAC ATA TTA GGC TCT GCC ATG CAA AAC ACA CAA AAT CTT CTC CAG ATG      2932
Asp Ile Leu Gly Ser Ala Met Gln Asn Thr Gln Asn Leu Leu Gln Met
        955                 960                 965

CCC TAT GGC TGT GGA GAG CAG AAT ATG GTC CTC TTT GCT CCT AAC ATC      2980
Pro Tyr Gly Cys Gly Glu Gln Asn Met Val Leu Phe Ala Pro Asn Ile
    970                 975                 980

TAT GTA CTG GAT TAT CTA AAT GAA ACA CAG CAG CTT ACT CCA GAG ATC      3028
Tyr Val Leu Asp Tyr Leu Asn Glu Thr Gln Gln Leu Thr Pro Glu Ile
985                 990                 995                1000

AAG TCC AAG GCC ATT GGC TAT CTC AAC ACT GGT TAC CAG AGA CAG TTG      3076
Lys Ser Lys Ala Ile Gly Tyr Leu Asn Thr Gly Tyr Gln Arg Gln Leu
                1005                1010                1015

AAC TAC AAA CAC TAT GAT GGC TCC TAC AGC ACC TTT GGG GAG CGA TAT      3124
Asn Tyr Lys His Tyr Asp Gly Ser Tyr Ser Thr Phe Gly Glu Arg Tyr
            1020                1025                1030

GGC AGG AAC CAG GGC AAC ACC TGG CTC ACA GCC TTT GTT CTG AAG ACT      3172
Gly Arg Asn Gln Gly Asn Thr Trp Leu Thr Ala Phe Val Leu Lys Thr
        1035                1040                1045

TTT GCC CAA GCT CGA GCC TAC ATC TTC ATC GAT GAA GCA CAC ATT ACC      3220
Phe Ala Gln Ala Arg Ala Tyr Ile Phe Ile Asp Glu Ala His Ile Thr
    1050                1055                1060

CAA GCC CTC ATA TGG CTC TCC CAG AGG CAG AAG GAC AAT GGC TGT TTC      3268
Gln Ala Leu Ile Trp Leu Ser Gln Arg Gln Lys Asp Asn Gly Cys Phe
1065                1070                1075                1080

AGG AGC TCT GGG TCA CTG CTC AAC AAT GCC ATA AAG GGA GGA GTA GAA      3316
Arg Ser Ser Gly Ser Leu Leu Asn Asn Ala Ile Lys Gly Gly Val Glu
                1085                1090                1095

GAT GAA GTG ACC CTC TCC GCC TAT ATC ACC ATC GCC CTT CTG GAG ATT      3364
Asp Glu Val Thr Leu Ser Ala Tyr Ile Thr Ile Ala Leu Leu Glu Ile
            1100                1105                1110

CCT CTC ACA GTC ACT CAC CCT GTT GTC CGC AAT GCC CTG TTT TGC CTG      3412
Pro Leu Thr Val Thr His Pro Val Val Arg Asn Ala Leu Phe Cys Leu
        1115                1120                1125

GAG TCA GCC TGG AAG ACA GCA CAA GAA GGG GAC CAT GGC AGC CAT GTA      3460
Glu Ser Ala Trp Lys Thr Ala Gln Glu Gly Asp His Gly Ser His Val
    1130                1135                1140
```

```
TAT ACC AAA GCA CTG CTG GCC TAT GCT TTT GCC CTG GCA GGT AAC CAG      3508
Tyr Thr Lys Ala Leu Leu Ala Tyr Ala Phe Ala Leu Ala Gly Asn Gln
1145                 1150                 1155                 1160

GAC AAG AGG AAG GAA GTA CTC AAG TCA CTT AAT GAG GAA GCT GTG AAG      3556
Asp Lys Arg Lys Glu Val Leu Lys Ser Leu Asn Glu Glu Ala Val Lys
                    1165                 1170                 1175

AAA GAC AAC TCT GTC CAT TGG GAG CGC CCT CAG AAA CCC AAG GCA CCA      3604
Lys Asp Asn Ser Val His Trp Glu Arg Pro Gln Lys Pro Lys Ala Pro
            1180                 1185                 1190

GTG GGG CAT TTT TAC GAA CCC CAG GCT CCC TCT GCT GAG GTG GAG ATG      3652
Val Gly His Phe Tyr Glu Pro Gln Ala Pro Ser Ala Glu Val Glu Met
        1195                 1200                 1205

ACA TCC TAT GTG CTC CTC GCT TAT CTC ACG GCC CAG CCA GCC CCA ACC      3700
Thr Ser Tyr Val Leu Leu Ala Tyr Leu Thr Ala Gln Pro Ala Pro Thr
    1210                 1215                 1220

TCG GAG GAC CTG ACC TCT GCA ACC AAC ATC GTG AAG TGG ATC ACG AAG      3748
Ser Glu Asp Leu Thr Ser Ala Thr Asn Ile Val Lys Trp Ile Thr Lys
1225                 1230                 1235                 1240

CAG CAG AAT GCC CAG GGC GGT TTC TCC TCC ACC CAG CAC ACA GTG GTG      3796
Gln Gln Asn Ala Gln Gly Gly Phe Ser Ser Thr Gln His Thr Val Val
                    1245                 1250                 1255

GCT CTC CAT GCT CTG TCC AAA TAT GGA GCA GCC ACA TTT ACC AGG ACT      3844
Ala Leu His Ala Leu Ser Lys Tyr Gly Ala Ala Thr Phe Thr Arg Thr
            1260                 1265                 1270

GGG AAG GCT GCA CAG GTG ACT ATC CAG TCT TCA GGG ACA TTT TCC AGC      3892
Gly Lys Ala Ala Gln Val Thr Ile Gln Ser Ser Gly Thr Phe Ser Ser
        1275                 1280                 1285

AAA TTC CAA GTG GAC AAC AAC AAC CGC CTG TTA CTG CAG CAG GTC TCA      3940
Lys Phe Gln Val Asp Asn Asn Asn Arg Leu Leu Leu Gln Gln Val Ser
    1290                 1295                 1300

TTG CCA GAG CTG CCT GGG GAA TAC AGC ATG AAA GTG ACA GGA GAA GGA      3988
Leu Pro Glu Leu Pro Gly Glu Tyr Ser Met Lys Val Thr Gly Glu Gly
1305                 1310                 1315                 1320

TGT GTC TAC CTC CAG ACA TCC TTG AAA TAC AAT ATT CTC CCA GAA AAG      4036
Cys Val Tyr Leu Gln Thr Ser Leu Lys Tyr Asn Ile Leu Pro Glu Lys
                    1325                 1330                 1335

GAA GAG TTC CCC TTT GCT TTA GGA GTG CAG ACT CTG CCT CAA ACT TGT      4084
Glu Glu Phe Pro Phe Ala Leu Gly Val Gln Thr Leu Pro Gln Thr Cys
            1340                 1345                 1350

GAT GAA CCC AAA GCC CAC ACC AGC TTC CAA ATC TCC CTA AGT GTC AGT      4132
Asp Glu Pro Lys Ala His Thr Ser Phe Gln Ile Ser Leu Ser Val Ser
        1355                 1360                 1365
```

```
TAC ACA GGG AGC CGC TCT GCC TCC AAC ATG GCG ATC GTT GAT GTG AAG      4180
Tyr Thr Gly Ser Arg Ser Ala Ser Asn Met Ala Ile Val Asp Val Lys
    1370                1375                1380

ATG GTC TCT GGC TTC ATT CCC CTG AAG CCA ACA GTG AAA ATG CTT GAA      4228
Met Val Ser Gly Phe Ile Pro Leu Lys Pro Thr Val Lys Met Leu Glu
1385                1390                1395                1400

AGA TCT AAC CAT GTG AGC CGG ACA GAA GTC AGC AGC AAC CAT GTC TTG      4276
Arg Ser Asn His Val Ser Arg Thr Glu Val Ser Ser Asn His Val Leu
                1405                1410                1415

ATT TAC CTT GAT AAG GTG TCA AAT CAG ACA CTG AGC TTG TTC TTC ACG      4324
Ile Tyr Leu Asp Lys Val Ser Asn Gln Thr Leu Ser Leu Phe Phe Thr
            1420                1425                1430

GTT CTG CAA GAT GTC CCA GTA AGA GAT CTC AAA CCA GCC ATA GTG AAA      4372
Val Leu Gln Asp Val Pro Val Arg Asp Leu Lys Pro Ala Ile Val Lys
        1435                1440                1445

GTC TAT GAT TAC TAC GAG ACG GAT GAG TTT GCA ATT GCT GAG TAC AAT      4420
Val Tyr Asp Tyr Tyr Glu Thr Asp Glu Phe Ala Ile Ala Glu Tyr Asn
    1450                1455                1460

GCT CCT TGC AGC AAA GAT CTT GGA AAT GCT TGAAGACCAC AAGGCTGAAA        4470
Ala Pro Cys Ser Lys Asp Leu Gly Asn Ala
1465                1470

AGTGCTTTGC TGGAGTCCTG TTCTCTGAGC TCCACAGAAG ACACGTGTTT TTGTATCTTT   4530

AAAGACTTGA TGAATAAACA CTTTTTCTGG TCAAAAAAA                          4569
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1474 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(E) FEATURES: bait region: 690-730
(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Gly Lys Asn Lys Leu Leu His Pro Ser Leu Val Leu Leu Leu Leu
 1               5                  10                  15

Val Leu Leu Pro Thr Asp Ala Ser Val Ser Gly Lys Pro Gln Tyr Met
            20                  25                  30

Val Leu Val Pro Ser Leu Leu His Thr Glu Thr Thr Glu Lys Gly Cys
        35                  40                  45

Val Leu Leu Ser Tyr Leu Asn Glu Thr Val Thr Val Ser Ala Ser Leu
    50                  55                  60

Glu Ser Val Arg Gly Asn Arg Ser Leu Phe Thr Asp Leu Glu Ala Glu
65                  70                  75                  80
```

Asn Asp Val Leu His Cys Val Ala Phe Ala Val Pro Lys Ser Ser Ser
85 90 95

Asn Glu Glu Val Met Phe Leu Thr Val Gln Val Lys Gly Pro Thr Gln
100 105 110

Glu Phe Lys Lys Arg Thr Thr Val Met Val Lys Asn Glu Asp Ser Leu
115 120 125

Val Phe Val Gln Thr Asp Lys Ser Ile Tyr Lys Pro Gly Gln Thr Val
130 135 140

Lys Phe Arg Val Val Ser Met Asp Glu Asn Phe His Pro Leu Asn Glu
145 150 155 160

Leu Ile Pro Leu Val Tyr Ile Gln Asp Pro Lys Gly Asn Arg Ile Ala
165 170 175

Gln Trp Gln Ser Phe Gln Leu Glu Gly Gly Leu Lys Gln Phe Ser Phe
180 185 190

Pro Leu Ser Ser Glu Pro Phe Gln Gly Ser Tyr Lys Val Val Val Gln
195 200 205

Lys Lys Ser Gly Gly Arg Thr Glu His Pro Phe Thr Val Glu Glu Phe
210 215 220

Val Leu Pro Lys Phe Glu Val Gln Val Thr Val Pro Lys Ile Ile Thr
225 230 235 240

Ile Leu Glu Glu Glu Met Asn Val Ser Val Cys Gly Leu Tyr Thr Tyr
245 250 255

Gly Lys Pro Val Pro Gly His Val Thr Val Ser Ile Cys Arg Lys Tyr
260 265 270

Ser Asp Ala Ser Asp Cys His Gly Glu Asp Ser Gln Ala Phe Cys Glu
275 280 285

Lys Phe Ser Gly Gln Leu Asn Ser His Gly Cys Phe Tyr Gln Gln Val
290 295 300

Lys Thr Lys Val Phe Gln Leu Lys Arg Lys Glu Tyr Glu Met Lys Leu
305 310 315 320

His Thr Glu Ala Gln Ile Gln Glu Glu Gly Thr Val Val Glu Leu Thr
325 330 335

Gly Arg Gln Ser Ser Glu Ile Thr Arg Thr Ile Thr Lys Leu Ser Phe
340 345 350

Val Lys Val Asp Ser His Phe Arg Gln Gly Ile Pro Phe Phe Gly Gln
355 360 365

Val Arg Leu Val Asp Gly Lys Gly Val Pro Ile Pro Asn Lys Val Ile
370 375 380

Phe Ile Arg Gly Asn Glu Ala Asn Tyr Tyr Ser Asn Ala Thr Thr Asp
385 390 395 400

Glu His Gly Leu Val Gln Phe Ser Ile Asn Thr Thr Asn Val Met Gly
405 410 415

Thr Ser Leu Thr Val Arg Val Asn Tyr Lys Asp Arg Ser Pro Cys Tyr
420 425 430

Gly Tyr Gln Trp Val Ser Glu Glu His Glu Glu Ala His His Thr Ala
435 440 445

Tyr Leu Val Phe Ser Pro Ser Lys Ser Phe Val His Leu Glu Pro Met
450 455 460

Ser His Glu Leu Pro Cys Gly His Thr Gln Thr Val Gln Ala His Tyr
465 470 475 480

Ile Leu Asn Gly Gly Thr Leu Leu Gly Leu Lys Lys Leu Ser Phe Tyr
485 490 495

Tyr Leu Ile Met Ala Lys Gly Gly Ile Val Arg Thr Gly Thr His Gly
500 505 510

Leu Leu Val Lys Gln Glu Asp Met Lys Gly His Phe Ser Ile Ser Ile
515 520 525

Pro Val Lys Ser Asp Ile Ala Pro Val Ala Arg Leu Leu Ile Tyr Ala
530 535 540

Val Leu Pro Thr Gly Asp Val Ile Gly Asp Ser Ala Lys Tyr Asp Val
545 550 555 560

Glu Asn Cys Leu Ala Asn Lys Val Asp Leu Ser Phe Ser Pro Ser Gln
565 570 575

Ser Leu Pro Ala Ser His Ala His Leu Arg Val Thr Ala Ala Pro Gln
580 585 590

Ser Val Cys Ala Leu Arg Ala Val Asp Gln Ser Val Leu Leu Met Lys
595 600 605

Pro Asp Ala Glu Leu Ser Ala Ser Ser Val Tyr Asn Leu Leu Pro Glu
610 615 620

Lys Asp Leu Thr Gly Phe Pro Gly Pro Leu Asn Asp Gln Asp Asp Glu
625 630 635 640

Asp Cys Ile Asn Arg His Asn Val Tyr Ile Asn Gly Ile Thr Tyr Thr
645 650 655

Pro Val Ser Ser Thr Asn Glu Lys Asp Met Tyr Ser Phe Leu Glu Asp
660 665 670

Met Gly Leu Lys Ala Phe Thr Asn Ser Lys Ile Arg Lys Pro Lys Met
675 680 685

Cys Pro Gln Leu Gln Gln Tyr Glu Met His Gly Pro Glu Gly Leu Arg
690 695 700

Val Gly Phe Tyr Glu Ser Asp Val Met Gly Arg Gly His Ala Arg Leu
705 710 715 720

Val His Val Glu Glu Pro His Thr Glu Thr Val Arg Lys Tyr Phe Pro
725 730 735

Glu Thr Trp Ile Trp Asp Leu Val Val Val Asn Ser Ala Gly Val Ala
740 745 750

Glu Val Gly Val Thr Val Pro Asp Thr Ile Thr Glu Trp Lys Ala Gly
755 760 765

Ala Phe Cys Leu Ser Glu Asp Ala Gly Leu Gly Ile Ser Ser Thr Ala
770 775 780

Ser Leu Arg Ala Phe Gln Pro Phe Phe Val Glu Leu Thr Met Pro Tyr
785 790 795 800

Ser Val Ile Arg Gly Glu Ala Phe Thr Leu Lys Ala Thr Val Leu Asn
805 810 815

Tyr Leu Pro Lys Cys Ile Arg Val Ser Val Gln Leu Glu Ala Ser Pro
820 825 830

Ala Phe Leu Ala Val Pro Val Glu Lys Glu Gln Ala Pro His Cys Ile
835 840 845

Cys Ala Asn Gly Arg Gln Thr Val Ser Trp Ala Val Thr Pro Lys Ser
850 855 860

Leu Gly Asn Val Asn Phe Thr Val Ser Ala Glu Ala Leu Glu Ser Gln
865 870 875 880

Glu Leu Cys Gly Thr Glu Val Pro Ser Val Pro Glu His Gly Arg Lys
885 890 895

Asp Thr Val Ile Lys Pro Leu Leu Val Glu Pro Glu Gly Leu Glu Lys
900 905 910

Glu Thr Thr Phe Asn Ser Leu Leu Cys Pro Ser Gly Gly Glu Val Ser
915 920 925

Glu Glu Leu Ser Leu Lys Leu Pro Pro Asn Val Val Glu Glu Ser Ala
930 935 940

Arg Ala Ser Val Ser Val Leu Gly Asp Ile Leu Gly Ser Ala Met Gln
945 950 955 960

Asn Thr Gln Asn Leu Leu Gln Met Pro Tyr Gly Cys Gly Glu Gln Asn
965 970 975

Met Val Leu Phe Ala Pro Asn Ile Tyr Val Leu Asp Tyr Leu Asn Glu
980 985 990

Thr Gln Gln Leu Thr Pro Glu Ile Lys Ser Lys Ala Ile Gly Tyr Leu
995 1000 1005

Asn Thr Gly Tyr Gln Arg Gln Leu Asn Tyr Lys His Tyr Asp Gly Ser
1010 1015 1020

Tyr Ser Thr Phe Gly Glu Arg Tyr Gly Arg Asn Gln Gly Asn Thr Trp
1025 1030 1035 1040

Leu Thr Ala Phe Val Leu Lys Thr Phe Ala Gln Ala Arg Ala Tyr Ile
1045 1050 1055

Phe Ile Asp Glu Ala His Ile Thr Gln Ala Leu Ile Trp Leu Ser Gln
1060 1065 1070

Arg Gln Lys Asp Asn Gly Cys Phe Arg Ser Ser Gly Ser Leu Leu Asn
1075 1080 1085

Asn Ala Ile Lys Gly Gly Val Glu Asp Glu Val Thr Leu Ser Ala Tyr
1090 1095 1100

Ile Thr Ile Ala Leu Leu Glu Ile Pro Leu Thr Val Thr His Pro Val
1105 1110 1115 1120

Val Arg Asn Ala Leu Phe Cys Leu Glu Ser Ala Trp Lys Thr Ala Gln
1125 1130 1135

Glu Gly Asp His Gly Ser His Val Tyr Thr Lys Ala Leu Leu Ala Tyr
1140 1145 1150

Ala Phe Ala Leu Ala Gly Asn Gln Asp Lys Arg Lys Glu Val Leu Lys
1155 1160 1165

Ser Leu Asn Glu Glu Ala Val Lys Lys Asp Asn Ser Val His Trp Glu
1170 1175 1180

Arg Pro Gln Lys Pro Lys Ala Pro Val Gly His Phe Tyr Glu Pro Gln
1185 1190 1195 1200

Ala Pro Ser Ala Glu Val Glu Met Thr Ser Tyr Val Leu Leu Ala Tyr
1205 1210 1215

Leu Thr Ala Gln Pro Ala Pro Thr Ser Glu Asp Leu Thr Ser Ala Thr
1220 1225 1230

Asn Ile Val Lys Trp Ile Thr Lys Gln Gln Asn Ala Gln Gly Gly Phe
1235 1240 1245

Ser Ser Thr Gln His Thr Val Val Ala Leu His Ala Leu Ser Lys Tyr
1250 1255 1260

Gly Ala Ala Thr Phe Thr Arg Thr Gly Lys Ala Ala Gln Val Thr Ile
1265 1270 1275 1280

Gln Ser Ser Gly Thr Phe Ser Ser Lys Phe Gln Val Asp Asn Asn Asn
1285 1290 1295

Arg Leu Leu Leu Gln Gln Val Ser Leu Pro Glu Leu Pro Gly Glu Tyr
1300 1305 1310

Ser Met Lys Val Thr Gly Glu Gly Cys Val Tyr Leu Gln Thr Ser Leu
1315 1320 1325

Lys Tyr Asn Ile Leu Pro Glu Lys Glu Glu Phe Pro Phe Ala Leu Gly
1330 1335 1340

Val Gln Thr Leu Pro Gln Thr Cys Asp Glu Pro Lys Ala His Thr Ser
1345 1350 1355 1360

Phe Gln Ile Ser Leu Ser Val Ser Tyr Thr Gly Ser Arg Ser Ala Ser
1365 1370 1375

Asn Met Ala Ile Val Asp Val Lys Met Val Ser Gly Phe Ile Pro Leu
1380 1385 1390

Lys Pro Thr Val Lys Met Leu Glu Arg Ser Asn His Val Ser Arg Thr
1395 1400 1405

Glu Val Ser Ser Asn His Val Leu Ile Tyr Leu Asp Lys Val Ser Asn
1410 1415 1420

Gln Thr Leu Ser Leu Phe Phe Thr Val Leu Gln Asp Val Pro Val Arg
1425 1430 1435 1440

Asp Leu Lys Pro Ala Ile Val Lys Val Tyr Asp Tyr Tyr Glu Thr Asp
1445 1450 1455

Glu Phe Ala Ile Ala Glu Tyr Asn Ala Pro Cys Ser Lys Asp Leu Gly
1460 1465 1470

Asn Ala

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4599 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: Y

(iv) ANTI-SENSE: N

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 29..4480
(D) OTHER INFORMATION:

(ix) FEATURE:
(A) NAME/KEY: insertion_seq
(B) LOCATION: 2102..2305
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
GTCTCCTCCA GCTCCTTCTT TCTGCAAC ATG GGG AAG AAC AAA CTC CTT CAT          52
                               Met Gly Lys Asn Lys Leu Leu His
                                 1               5

CCA AGT CTG GTT CTT CTC CTC TTG GTC CTC CTG CCC ACA GAC GCC TCA        100
Pro Ser Leu Val Leu Leu Leu Leu Val Leu Leu Pro Thr Asp Ala Ser
     10                  15                  20

GTC TCT GGA AAA CCG CAG TAT ATG GTT CTG GTC CCC TCC CTG CTC CAC        148
Val Ser Gly Lys Pro Gln Tyr Met Val Leu Val Pro Ser Leu Leu His
 25                  30                  35                  40

ACT GAG ACC ACT GAG AAG GGC TGT GTC CTT CTG AGC TAC CTG AAT GAG        196
Thr Glu Thr Thr Glu Lys Gly Cys Val Leu Leu Ser Tyr Leu Asn Glu
                    45                  50                  55

ACA GTG ACT GTA AGT GCT TCC TTG GAG TCT GTC AGG GGA AAC AGG AGC        244
Thr Val Thr Val Ser Ala Ser Leu Glu Ser Val Arg Gly Asn Arg Ser
                60                  65                  70

CTC TTC ACT GAC CTG GAG GCG GAG AAT GAC GTA CTC CAC TGT GTC GCC        292
Leu Phe Thr Asp Leu Glu Ala Glu Asn Asp Val Leu His Cys Val Ala
          75                  80                  85

TTC GCT GTC CCA AAG TCT TCA TCC AAT GAG GAG GTA ATG TTC CTC ACT        340
Phe Ala Val Pro Lys Ser Ser Ser Asn Glu Glu Val Met Phe Leu Thr
     90                  95                  100

GTC CAA GTG AAA GGA CCA ACC CAA GAA TTT AAG AAG CGG ACC ACA GTG        388
Val Gln Val Lys Gly Pro Thr Gln Glu Phe Lys Lys Arg Thr Thr Val
105                 110                 115                 120

ATG GTT AAG AAC GAG GAC AGT CTG GTC TTT GTC CAG ACA GAC AAA TCA        436
Met Val Lys Asn Glu Asp Ser Leu Val Phe Val Gln Thr Asp Lys Ser
                125                 130                 135

ATC TAC AAA CCA GGG CAG ACA GTG AAA TTT CGT GTT GTC TCC ATG GAT        484
Ile Tyr Lys Pro Gly Gln Thr Val Lys Phe Arg Val Val Ser Met Asp
           140                 145                 150

GAA AAC TTT CAC CCC CTG AAT GAG TTG ATT CCA CTA GTA TAC ATT CAG        532
Glu Asn Phe His Pro Leu Asn Glu Leu Ile Pro Leu Val Tyr Ile Gln
      155                 160                 165

GAT CCC AAA GGA AAT CGC ATC GCA CAA TGG CAG AGT TTC CAG TTA GAG        580
Asp Pro Lys Gly Asn Arg Ile Ala Gln Trp Gln Ser Phe Gln Leu Glu
 170                 175                 180
```

```
GGT GGC CTC AAG CAA TTT TCT TTT CCC CTC TCA TCA GAG CCC TTC CAG          628
Gly Gly Leu Lys Gln Phe Ser Phe Pro Leu Ser Ser Glu Pro Phe Gln
185                 190                 195                 200

GGC TCC TAC AAG GTG GTG GTA CAG AAG AAA TCA GGT GGA AGG ACA GAG          676
Gly Ser Tyr Lys Val Val Val Gln Lys Lys Ser Gly Gly Arg Thr Glu
                205                 210                 215

CAC CCT TTC ACC GTG GAG GAA TTT GTT CTT CCC AAG TTT GAA GTA CAA          724
His Pro Phe Thr Val Glu Glu Phe Val Leu Pro Lys Phe Glu Val Gln
            220                 225                 230

GTA ACA GTG CCA AAG ATA ATC ACC ATC TTG GAA GAA GAG ATG AAT GTA          772
Val Thr Val Pro Lys Ile Ile Thr Ile Leu Glu Glu Glu Met Asn Val
        235                 240                 245

TCA GTG TGT GGC CTA TAC ACA TAT GGG AAG CCT GTC CCT GGA CAT GTG          820
Ser Val Cys Gly Leu Tyr Thr Tyr Gly Lys Pro Val Pro Gly His Val
    250                 255                 260

ACT GTG AGC ATT TGC AGA AAG TAT AGT GAC GCT TCC GAC TGC CAC GGT          868
Thr Val Ser Ile Cys Arg Lys Tyr Ser Asp Ala Ser Asp Cys His Gly
265                 270                 275                 280

GAA GAT TCA CAG GCT TTC TGT GAG AAA TTC AGT GGA CAG CTA AAC AGC          916
Glu Asp Ser Gln Ala Phe Cys Glu Lys Phe Ser Gly Gln Leu Asn Ser
                285                 290                 295

CAT GGC TGC TTC TAT CAG CAA GTA AAA ACC AAG GTC TTC CAG CTG AAG          964
His Gly Cys Phe Tyr Gln Gln Val Lys Thr Lys Val Phe Gln Leu Lys
            300                 305                 310

AGG AAG GAG TAT GAA ATG AAA CTT CAC ACT GAG GCC CAG ATC CAA GAA         1012
Arg Lys Glu Tyr Glu Met Lys Leu His Thr Glu Ala Gln Ile Gln Glu
        315                 320                 325

GAA GGA ACA GTG GTG GAA TTG ACT GGA AGG CAG TCC AGT GAA ATC ACA         1060
Glu Gly Thr Val Val Glu Leu Thr Gly Arg Gln Ser Ser Glu Ile Thr
    330                 335                 340

AGA ACC ATA ACC AAA CTC TCA TTT GTG AAA GTG GAC TCA CAC TTT CGA         1108
Arg Thr Ile Thr Lys Leu Ser Phe Val Lys Val Asp Ser His Phe Arg
345                 350                 355                 360

CAG GGA ATT CCC TTC TTT GGG CAG GTG CGC CTA GTA GAT GGG AAA GGC         1156
Gln Gly Ile Pro Phe Phe Gly Gln Val Arg Leu Val Asp Gly Lys Gly
                365                 370                 375

GTC CCT ATA CCA AAT AAA GTC ATA TTC ATC AGA GGA AAT GAA GCA AAC         1204
Val Pro Ile Pro Asn Lys Val Ile Phe Ile Arg Gly Asn Glu Ala Asn
            380                 385                 390

TAT TAC TCC AAT GCT ACC ACG GAT GAG CAT GGC CTT GTA CAG TTC TCT         1252
Tyr Tyr Ser Asn Ala Thr Thr Asp Glu His Gly Leu Val Gln Phe Ser
        395                 400                 405
```

```
ATC AAC ACC ACC AAT GTT ATG GGT ACC TCT CTT ACT GTT AGG GTC AAT      1300
Ile Asn Thr Thr Asn Val Met Gly Thr Ser Leu Thr Val Arg Val Asn
    410                 415                 420

TAC AAG GAT CGT AGT CCC TGT TAC GGC TAC CAG TGG GTG TCA GAA GAA      1348
Tyr Lys Asp Arg Ser Pro Cys Tyr Gly Tyr Gln Trp Val Ser Glu Glu
425                 430                 435                 440

CAC GAA GAG GCA CAT CAC ACT GCT TAT CTT GTG TTC TCC CCA AGC AAG      1396
His Glu Glu Ala His His Thr Ala Tyr Leu Val Phe Ser Pro Ser Lys
                445                 450                 455

AGC TTT GTC CAC CTT GAG CCC ATG TCT CAT GAA CTA CCC TGT GGC CAT      1444
Ser Phe Val His Leu Glu Pro Met Ser His Glu Leu Pro Cys Gly His
            460                 465                 470

ACT CAG ACA GTC CAG GCA CAT TAT ATT CTG AAT GGA GGC ACC CTG CTG      1492
Thr Gln Thr Val Gln Ala His Tyr Ile Leu Asn Gly Gly Thr Leu Leu
        475                 480                 485

GGG CTG AAG AAG CTC TCC TTC TAT TAT CTG ATA ATG GCA AAG GGA GGC      1540
Gly Leu Lys Lys Leu Ser Phe Tyr Tyr Leu Ile Met Ala Lys Gly Gly
    490                 495                 500

ATT GTC CGA ACT GGG ACT CAT GGA CTG CTT GTG AAG CAG GAA GAC ATG      1588
Ile Val Arg Thr Gly Thr His Gly Leu Leu Val Lys Gln Glu Asp Met
505                 510                 515                 520

AAG GGC CAT TTT TCC ATC TCA ATC CCT GTG AAG TCA GAC ATT GCT CCT      1636
Lys Gly His Phe Ser Ile Ser Ile Pro Val Lys Ser Asp Ile Ala Pro
                525                 530                 535

GTC GCT CGG TTG CTC ATC TAT GCT GTT TTA CCT ACC GGG GAC GTG ATT      1684
Val Ala Arg Leu Leu Ile Tyr Ala Val Leu Pro Thr Gly Asp Val Ile
            540                 545                 550

GGG GAT TCT GCA AAA TAT GAT GTT GAA AAT TGT CTG GCC AAC AAG GTG      1732
Gly Asp Ser Ala Lys Tyr Asp Val Glu Asn Cys Leu Ala Asn Lys Val
        555                 560                 565

GAT TTG AGC TTC AGC CCA TCA CAA AGT CTC CCA GCC TCA CAC GCC CAC      1780
Asp Leu Ser Phe Ser Pro Ser Gln Ser Leu Pro Ala Ser His Ala His
    570                 575                 580

CTG CGA GTC ACA GCG GCT CCT CAG TCC GTC TGC GCC CTC CGT GCT GTG      1828
Leu Arg Val Thr Ala Ala Pro Gln Ser Val Cys Ala Leu Arg Ala Val
585                 590                 595                 600

GAC CAA AGC GTG CTG CTC ATG AAG CCT GAT GCT GAG CTC TCG GCG TCC      1876
Asp Gln Ser Val Leu Leu Met Lys Pro Asp Ala Glu Leu Ser Ala Ser
                605                 610                 615

TCG GTT TAC AAC CTG CTA CCA GAA AAG GAC CTC ACT GGC TTC CCT GGG      1924
Ser Val Tyr Asn Leu Leu Pro Glu Lys Asp Leu Thr Gly Phe Pro Gly
            620                 625                 630
```

```
CCT TTG AAT GAC CAG GAC GAT GAA GAC TGC ATC AAT CGT CAT AAT GTC      1972
Pro Leu Asn Asp Gln Asp Asp Glu Asp Cys Ile Asn Arg His Asn Val
        635                 640                 645

TAT ATT AAT GGA ATC ACA TAT ACT CCA GTA TCA AGT ACA AAT GAA AAG      2020
Tyr Ile Asn Gly Ile Thr Tyr Thr Pro Val Ser Ser Thr Asn Glu Lys
    650                 655                 660

GAT ATG TAC AGC TTC CTA GAG GAC ATG GGC TTA AAG GCA TTC ACC AAC      2068
Asp Met Tyr Ser Phe Leu Glu Asp Met Gly Leu Lys Ala Phe Thr Asn
665                 670                 675                 680

TCA AAG ATT CGT AAA CCC AAA ATG TGT CCA CAG CTG CAG TCA GTG TCA      2116
Ser Lys Ile Arg Lys Pro Lys Met Cys Pro Gln Leu Gln Ser Val Ser
                685                 690                 695

GCC GGC GCC GTG GGA CAG GGA TAT TAT GGA GCC GGA CTG GGA GTG GTG      2164
Ala Gly Ala Val Gly Gln Gly Tyr Tyr Gly Ala Gly Leu Gly Val Val
            700                 705                 710

GAG AGG CCT TAT GTG CCT CAG CTG GGT ACC TAT AAT GTG ATC CCT CTG      2212
Glu Arg Pro Tyr Val Pro Gln Leu Gly Thr Tyr Asn Val Ile Pro Leu
        715                 720                 725

AAT AAT GAG CAG AGC TCA GGA CCT GTG CCT GAG ACA GTG AGG AAG TAT      2260
Asn Asn Glu Gln Ser Ser Gly Pro Val Pro Glu Thr Val Arg Lys Tyr
    730                 735                 740

TTC CCT GAG ACA TGG ATC TGG GAT CTG GTG GTG GTG AAT TCC GCG GGT      2308
Phe Pro Glu Thr Trp Ile Trp Asp Leu Val Val Val Asn Ser Ala Gly
745                 750                 755                 760

GTG GCT GAG GTA GGA GTA ACA GTC CCT GAC ACC ATC ACC GAG TGG AAG      2356
Val Ala Glu Val Gly Val Thr Val Pro Asp Thr Ile Thr Glu Trp Lys
                765                 770                 775

GCA GGG GCC TTC TGC CTG TCT GAA GAT GCT GGA CTT GGT ATC TCT TCC      2404
Ala Gly Ala Phe Cys Leu Ser Glu Asp Ala Gly Leu Gly Ile Ser Ser
            780                 785                 790

ACT GCC TCT CTC CGA GCC TTC CAG CCC TTC TTT GTG GAG CTC ACA ATG      2452
Thr Ala Ser Leu Arg Ala Phe Gln Pro Phe Phe Val Glu Leu Thr Met
        795                 800                 805

CCT TAC TCT GTG ATT CGT GGA GAG GCC TTC ACA CTC AAG GCC ACG GTC      2500
Pro Tyr Ser Val Ile Arg Gly Glu Ala Phe Thr Leu Lys Ala Thr Val
    810                 815                 820

CTA AAC TAC CTT CCC AAA TGC ATC CGG GTC AGT GTG CAG CTG GAA GCC      2548
Leu Asn Tyr Leu Pro Lys Cys Ile Arg Val Ser Val Gln Leu Glu Ala
825                 830                 835                 840

TCT CCC GCC TTC CTA GCT GTC CCA GTG GAG AAG GAA CAA GCG CCT CAC      2596
Ser Pro Ala Phe Leu Ala Val Pro Val Glu Lys Glu Gln Ala Pro His
                845                 850                 855
```

```
TGC ATC TGT GCA AAC GGG CGG CAA ACT GTG TCC TGG GCA GTA ACC CCA      2644
Cys Ile Cys Ala Asn Gly Arg Gln Thr Val Ser Trp Ala Val Thr Pro
            860                 865                 870

AAG TCA TTA GGA AAT GTG AAT TTC ACT GTG AGC GCA GAG GCA CTA GAG      2692
Lys Ser Leu Gly Asn Val Asn Phe Thr Val Ser Ala Glu Ala Leu Glu
        875                 880                 885

TCT CAA GAG CTG TGT GGG ACT GAG GTG CCT TCA GTT CCT GAA CAC GGA      2740
Ser Gln Glu Leu Cys Gly Thr Glu Val Pro Ser Val Pro Glu His Gly
    890                 895                 900

AGG AAA GAC ACA GTC ATC AAG CCT CTG TTG GTT GAA CCT GAA GGA CTA      2788
Arg Lys Asp Thr Val Ile Lys Pro Leu Leu Val Glu Pro Glu Gly Leu
905                 910                 915                 920

GAG AAG GAA ACA ACA TTC AAC TCC CTA CTT TGT CCA TCA GGT GGT GAG      2836
Glu Lys Glu Thr Thr Phe Asn Ser Leu Leu Cys Pro Ser Gly Gly Glu
                925                 930                 935

GTT TCT GAA GAA TTA TCC CTG AAA CTG CCA CCA AAT GTG GTA GAA GAA      2884
Val Ser Glu Glu Leu Ser Leu Lys Leu Pro Pro Asn Val Val Glu Glu
            940                 945                 950

TCT GCC CGA GCT TCT GTC TCA GTT TTG GGA GAC ATA TTA GGC TCT GCC      2932
Ser Ala Arg Ala Ser Val Ser Val Leu Gly Asp Ile Leu Gly Ser Ala
        955                 960                 965

ATG CAA AAC ACA CAA AAT CTT CTC CAG ATG CCC TAT GGC TGT GGA GAG      2980
Met Gln Asn Thr Gln Asn Leu Leu Gln Met Pro Tyr Gly Cys Gly Glu
    970                 975                 980

CAG AAT ATG GTC CTC TTT GCT CCT AAC ATC TAT GTA CTG GAT TAT CTA      3028
Gln Asn Met Val Leu Phe Ala Pro Asn Ile Tyr Val Leu Asp Tyr Leu
985                 990                 995                 1000

AAT GAA ACA CAG CAG CTT ACT CCA GAG ATC AAG TCC AAG GCC ATT GGC      3076
Asn Glu Thr Gln Gln Leu Thr Pro Glu Ile Lys Ser Lys Ala Ile Gly
                1005                1010                1015

TAT CTC AAC ACT GGT TAC CAG AGA CAG TTG AAC TAC AAA CAC TAT GAT      3124
Tyr Leu Asn Thr Gly Tyr Gln Arg Gln Leu Asn Tyr Lys His Tyr Asp
            1020                1025                1030

GGC TCC TAC AGC ACC TTT GGG GAG CGA TAT GGC AGG AAC CAG GGC AAC      3172
Gly Ser Tyr Ser Thr Phe Gly Glu Arg Tyr Gly Arg Asn Gln Gly Asn
        1035                1040                1045

ACC TGG CTC ACA GCC TTT GTT CTG AAG ACT TTT GCC CAA GCT CGA GCC      3220
Thr Trp Leu Thr Ala Phe Val Leu Lys Thr Phe Ala Gln Ala Arg Ala
    1050                1055                1060

TAC ATC TTC ATC GAT GAA GCA CAC ATT ACC CAA GCC CTC ATA TGG CTC      3268
Tyr Ile Phe Ile Asp Glu Ala His Ile Thr Gln Ala Leu Ile Trp Leu
1065                1070                1075                1080
```

```
TCC CAG AGG CAG AAG GAC AAT GGC TGT TTC AGG AGC TCT GGG TCA CTG      3316
Ser Gln Arg Gln Lys Asp Asn Gly Cys Phe Arg Ser Ser Gly Ser Leu
                1085                1090                1095

CTC AAC AAT GCC ATA AAG GGA GGA GTA GAA GAT GAA GTG ACC CTC TCC      3364
Leu Asn Asn Ala Ile Lys Gly Gly Val Glu Asp Glu Val Thr Leu Ser
            1100                1105                1110

GCC TAT ATC ACC ATC GCC CTT CTG GAG ATT CCT CTC ACA GTC ACT CAC      3412
Ala Tyr Ile Thr Ile Ala Leu Leu Glu Ile Pro Leu Thr Val Thr His
        1115                1120                1125

CCT GTT GTC CGC AAT GCC CTG TTT TGC CTG GAG TCA GCC TGG AAG ACA      3460
Pro Val Val Arg Asn Ala Leu Phe Cys Leu Glu Ser Ala Trp Lys Thr
    1130                1135                1140

GCA CAA GAA GGG GAC CAT GGC AGC CAT GTA TAT ACC AAA GCA CTG CTG      3508
Ala Gln Glu Gly Asp His Gly Ser His Val Tyr Thr Lys Ala Leu Leu
1145                1150                1155                1160

GCC TAT GCT TTT GCC CTG GCA GGT AAC CAG GAC AAG AGG AAG GAA GTA      3556
Ala Tyr Ala Phe Ala Leu Ala Gly Asn Gln Asp Lys Arg Lys Glu Val
                1165                1170                1175

CTC AAG TCA CTT AAT GAG GAA GCT GTG AAG AAA GAC AAC TCT GTC CAT      3604
Leu Lys Ser Leu Asn Glu Glu Ala Val Lys Lys Asp Asn Ser Val His
            1180                1185                1190

TGG GAG CGC CCT CAG AAA CCC AAG GCA CCA GTG GGG CAT TTT TAC GAA      3652
Trp Glu Arg Pro Gln Lys Pro Lys Ala Pro Val Gly His Phe Tyr Glu
        1195                1200                1205

CCC CAG GCT CCC TCT GCT GAG GTG GAG ATG ACA TCC TAT GTG CTC CTC      3700
Pro Gln Ala Pro Ser Ala Glu Val Glu Met Thr Ser Tyr Val Leu Leu
    1210                1215                1220

GCT TAT CTC ACG GCC CAG CCA GCC CCA ACC TCG GAG GAC CTG ACC TCT      3748
Ala Tyr Leu Thr Ala Gln Pro Ala Pro Thr Ser Glu Asp Leu Thr Ser
1225                1230                1235                1240

GCA ACC AAC ATC GTG AAG TGG ATC ACG AAG CAG CAG AAT GCC CAG GGC      3796
Ala Thr Asn Ile Val Lys Trp Ile Thr Lys Gln Gln Asn Ala Gln Gly
                1245                1250                1255

GGT TTC TCC TCC ACC CAG CAC ACA GTG GTG GCT CTC CAT GCT CTG TCC      3844
Gly Phe Ser Ser Thr Gln His Thr Val Val Ala Leu His Ala Leu Ser
            1260                1265                1270

AAA TAT GGA GCA GCC ACA TTT ACC AGG ACT GGG AAG GCT GCA CAG GTG      3892
Lys Tyr Gly Ala Ala Thr Phe Thr Arg Thr Gly Lys Ala Ala Gln Val
        1275                1280                1285

ACT ATC CAG TCT TCA GGG ACA TTT TCC AGC AAA TTC CAA GTG GAC AAC      3940
Thr Ile Gln Ser Ser Gly Thr Phe Ser Ser Lys Phe Gln Val Asp Asn
    1290                1295                1300
```

```
AAC AAC CGC CTG TTA CTG CAG CAG GTC TCA TTG CCA GAG CTG CCT GGG      3988
Asn Asn Arg Leu Leu Leu Gln Gln Val Ser Leu Pro Glu Leu Pro Gly
1305                1310                1315                1320

GAA TAC AGC ATG AAA GTG ACA GGA GAA GGA TGT GTC TAC CTC CAG ACA      4036
Glu Tyr Ser Met Lys Val Thr Gly Glu Gly Cys Val Tyr Leu Gln Thr
                1325                1330                1335

TCC TTG AAA TAC AAT ATT CTC CCA GAA AAG GAA GAG TTC CCC TTT GCT      4084
Ser Leu Lys Tyr Asn Ile Leu Pro Glu Lys Glu Glu Phe Pro Phe Ala
            1340                1345                1350

TTA GGA GTG CAG ACT CTG CCT CAA ACT TGT GAT GAA CCC AAA GCC CAC      4132
Leu Gly Val Gln Thr Leu Pro Gln Thr Cys Asp Glu Pro Lys Ala His
        1355                1360                1365

ACC AGC TTC CAA ATC TCC CTA AGT GTC AGT TAC ACA GGG AGC CGC TCT      4180
Thr Ser Phe Gln Ile Ser Leu Ser Val Ser Tyr Thr Gly Ser Arg Ser
    1370                1375                1380

GCC TCC AAC ATG GCG ATC GTT GAT GTG AAG ATG GTC TCT GGC TTC ATT      4228
Ala Ser Asn Met Ala Ile Val Asp Val Lys Met Val Ser Gly Phe Ile
1385                1390                1395                1400

CCC CTG AAG CCA ACA GTG AAA ATG CTT GAA AGA TCT AAC CAT GTG AGC      4276
Pro Leu Lys Pro Thr Val Lys Met Leu Glu Arg Ser Asn His Val Ser
                1405                1410                1415

CGG ACA GAA GTC AGC AGC AAC CAT GTC TTG ATT TAC CTT GAT AAG GTG      4324
Arg Thr Glu Val Ser Ser Asn His Val Leu Ile Tyr Leu Asp Lys Val
            1420                1425                1430

TCA AAT CAG ACA CTG AGC TTG TTC TTC ACG GTT CTG CAA GAT GTC CCA      4372
Ser Asn Gln Thr Leu Ser Leu Phe Phe Thr Val Leu Gln Asp Val Pro
        1435                1440                1445

GTA AGA GAT CTG AAA CCA GCC ATA GTG AAA GTC TAT GAT TAC TAC GAG      4420
Val Arg Asp Leu Lys Pro Ala Ile Val Lys Val Tyr Asp Tyr Tyr Glu
    1450                1455                1460

ACG GAT GAG TTT GCA ATT GCT GAG TAC AAT GCT CCT TGC AGC AAA GAT      4468
Thr Asp Glu Phe Ala Ile Ala Glu Tyr Asn Ala Pro Cys Ser Lys Asp
1465                1470                1475                1480

CTT GGA AAT GCT TGAAGACCAC AAGGCTGAAA AGTGCTTTGC TGGAGTCCTG          4520
Leu Gly Asn Ala

TTCTCTGAGC TCCACAGAAG ACACGTGTTT TTGTATCTTT AAAGACTTGA TGAATAAACA    4580

CTTTTTCTGG TCAAAAAAA                                                4599
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1484 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear
(E) FEATURES: bait region: 690-740
(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Gly Lys Asn Lys Leu Leu His Pro Ser Leu Val Leu Leu Leu Leu
1 5 10 15

Val Leu Leu Pro Thr Asp Ala Ser Val Ser Gly Lys Pro Gln Tyr Met
20 25 30

Val Leu Val Pro Ser Leu Leu His Thr Glu Thr Thr Glu Lys Gly Cys
35 40 45

Val Leu Leu Ser Tyr Leu Asn Glu Thr Val Thr Val Ser Ala Ser Leu
50 55 60

Glu Ser Val Arg Gly Asn Arg Ser Leu Phe Thr Asp Leu Glu Ala Glu
65 70 75 80

Asn Asp Val Leu His Cys Val Ala Phe Ala Val Pro Lys Ser Ser Ser
85 90 95

Asn Glu Glu Val Met Phe Leu Thr Val Gln Val Lys Gly Pro Thr Gln
100 105 110

Glu Phe Lys Lys Arg Thr Thr Val Met Val Lys Asn Glu Asp Ser Leu
115 120 125

Val Phe Val Gln Thr Asp Lys Ser Ile Tyr Lys Pro Gly Gln Thr Val
130 135 140

Lys Phe Arg Val Val Ser Met Asp Glu Asn Phe His Pro Leu Asn Glu
145 150 155 160

Leu Ile Pro Leu Val Tyr Ile Gln Asp Pro Lys Gly Asn Arg Ile Ala
165 170 175

Gln Trp Gln Ser Phe Gln Leu Glu Gly Gly Leu Lys Gln Phe Ser Phe
180 185 190

Pro Leu Ser Ser Glu Pro Phe Gln Gly Ser Tyr Lys Val Val Val Gln
195 200 205

Lys Lys Ser Gly Gly Arg Thr Glu His Pro Phe Thr Val Glu Glu Phe
210 215 220

Val Leu Pro Lys Phe Glu Val Gln Val Thr Val Pro Lys Ile Ile Thr
225 230 235 240

Ile Leu Glu Glu Glu Met Asn Val Ser Val Cys Gly Leu Tyr Thr Tyr
245 250 255

Gly Lys Pro Val Pro Gly His Val Thr Val Ser Ile Cys Arg Lys Tyr
260 265 270

```
Ser Asp Ala Ser Asp Cys His Gly Glu Asp Ser Gln Ala Phe Cys Glu
        275                 280                 285

Lys Phe Ser Gly Gln Leu Asn Ser His Gly Cys Phe Tyr Gln Gln Val
    290                 295                 300

Lys Thr Lys Val Phe Gln Leu Lys Arg Lys Glu Tyr Glu Met Lys Leu
305                 310                 315                 320

His Thr Glu Ala Gln Ile Gln Glu Glu Gly Thr Val Val Glu Leu Thr
                325                 330                 335

Gly Arg Gln Ser Ser Glu Ile Thr Arg Thr Ile Thr Lys Leu Ser Phe
            340                 345                 350

Val Lys Val Asp Ser His Phe Arg Gln Gly Ile Pro Phe Phe Gly Gln
        355                 360                 365

Val Arg Leu Val Asp Gly Lys Gly Val Pro Ile Pro Asn Lys Val Ile
    370                 375                 380

Phe Ile Arg Gly Asn Glu Ala Asn Tyr Tyr Ser Asn Ala Thr Thr Asp
385                 390                 395                 400

Glu His Gly Leu Val Gln Phe Ser Ile Asn Thr Thr Asn Val Met Gly
                405                 410                 415

Thr Ser Leu Thr Val Arg Val Asn Tyr Lys Asp Arg Ser Pro Cys Tyr
            420                 425                 430

Gly Tyr Gln Trp Val Ser Glu Glu His Glu Glu Ala His His Thr Ala
        435                 440                 445

Tyr Leu Val Phe Ser Pro Ser Lys Ser Phe Val His Leu Glu Pro Met
    450                 455                 460

Ser His Glu Leu Pro Cys Gly His Thr Gln Thr Val Gln Ala His Tyr
465                 470                 475                 480

Ile Leu Asn Gly Gly Thr Leu Leu Gly Leu Lys Lys Leu Ser Phe Tyr
                485                 490                 495

Tyr Leu Ile Met Ala Lys Gly Gly Ile Val Arg Thr Gly Thr His Gly
            500                 505                 510

Leu Leu Val Lys Gln Glu Asp Met Lys Gly His Phe Ser Ile Ser Ile
        515                 520                 525

Pro Val Lys Ser Asp Ile Ala Pro Val Ala Arg Leu Leu Ile Tyr Ala
    530                 535                 540

Val Leu Pro Thr Gly Asp Val Ile Gly Asp Ser Ala Lys Tyr Asp Val
545                 550                 555                 560

Glu Asn Cys Leu Ala Asn Lys Val Asp Leu Ser Phe Ser Pro Ser Gln
                565                 570                 575
```

Ser Leu Pro Ala Ser His Ala His Leu Arg Val Thr Ala Ala Pro Gln
580 585 590

Ser Val Cys Ala Leu Arg Ala Val Asp Gln Ser Val Leu Leu Met Lys
595 600 605

Pro Asp Ala Glu Leu Ser Ala Ser Ser Val Tyr Asn Leu Leu Pro Glu
610 615 620

Lys Asp Leu Thr Gly Phe Pro Gly Pro Leu Asn Asp Gln Asp Asp Glu
625 630 635 640

Asp Cys Ile Asn Arg His Asn Val Tyr Ile Asn Gly Ile Thr Tyr Thr
645 650 655

Pro Val Ser Ser Thr Asn Glu Lys Asp Met Tyr Ser Phe Leu Glu Asp
660 665 670

Met Gly Leu Lys Ala Phe Thr Asn Ser Lys Ile Arg Lys Pro Lys Met
675 680 685

Cys Pro Gln Leu Gln Ser Val Ser Ala Gly Ala Val Gly Gln Gly Tyr
690 695 700

Tyr Gly Ala Gly Leu Gly Val Val Glu Arg Pro Tyr Val Pro Gln Leu
705 710 715 720

Gly Thr Tyr Asn Val Ile Pro Leu Asn Asn Glu Gln Ser Ser Gly Pro
725 730 735

Val Pro Glu Thr Val Arg Lys Tyr Phe Pro Glu Thr Trp Ile Trp Asp
740 745 750

Leu Val Val Val Asn Ser Ala Gly Val Ala Glu Val Gly Val Thr Val
755 760 765

Pro Asp Thr Ile Thr Glu Trp Lys Ala Gly Ala Phe Cys Leu Ser Glu
770 775 780

Asp Ala Gly Leu Gly Ile Ser Ser Thr Ala Ser Leu Arg Ala Phe Gln
785 790 795 800

Pro Phe Phe Val Glu Leu Thr Met Pro Tyr Ser Val Ile Arg Gly Glu
805 810 815

Ala Phe Thr Leu Lys Ala Thr Val Leu Asn Tyr Leu Pro Lys Cys Ile
820 825 830

Arg Val Ser Val Gln Leu Glu Ala Ser Pro Ala Phe Leu Ala Val Pro
835 840 845

Val Glu Lys Glu Gln Ala Pro His Cys Ile Cys Ala Asn Gly Arg Gln
850 855 860

Thr Val Ser Trp Ala Val Thr Pro Lys Ser Leu Gly Asn Val Asn Phe
865 870 875 880

Thr Val Ser Ala Glu Ala Leu Glu Ser Gln Glu Leu Cys Gly Thr Glu
885 890 895

Val Pro Ser Val Pro Glu His Gly Arg Lys Asp Thr Val Ile Lys Pro
900 905 910

Leu Leu Val Glu Pro Glu Gly Leu Glu Lys Glu Thr Thr Phe Asn Ser
915 920 925

Leu Leu Cys Pro Ser Gly Gly Glu Val Ser Glu Glu Leu Ser Leu Lys
930 935 940

Leu Pro Pro Asn Val Val Glu Glu Ser Ala Arg Ala Ser Val Ser Val
945 950 955 960

Leu Gly Asp Ile Leu Gly Ser Ala Met Gln Asn Thr Gln Asn Leu Leu
965 970 975

Gln Met Pro Tyr Gly Cys Gly Glu Gln Asn Met Val Leu Phe Ala Pro
980 985 990

Asn Ile Tyr Val Leu Asp Tyr Leu Asn Glu Thr Gln Gln Leu Thr Pro
995 1000 1005

Glu Ile Lys Ser Lys Ala Ile Gly Tyr Leu Asn Thr Gly Tyr Gln Arg
1010 1015 1020

Gln Leu Asn Tyr Lys His Tyr Asp Gly Ser Tyr Ser Thr Phe Gly Glu
1025 1030 1035 1040

Arg Tyr Gly Arg Asn Gln Gly Asn Thr Trp Leu Thr Ala Phe Val Leu
1045 1050 1055

Lys Thr Phe Ala Gln Ala Arg Ala Tyr Ile Phe Ile Asp Glu Ala His
1060 1065 1070

Ile Thr Gln Ala Leu Ile Trp Leu Ser Gln Arg Gln Lys Asp Asn Gly
1075 1080 1085

Cys Phe Arg Ser Ser Gly Ser Leu Leu Asn Asn Ala Ile Lys Gly Gly
1090 1095 1100

Val Glu Asp Glu Val Thr Leu Ser Ala Tyr Ile Thr Ile Ala Leu Leu
1105 1110 1115 1120

Glu Ile Pro Leu Thr Val Thr His Pro Val Val Arg Asn Ala Leu Phe
1125 1130 1135

Cys Leu Glu Ser Ala Trp Lys Thr Ala Gln Glu Gly Asp His Gly Ser
1140 1145 1150

His Val Tyr Thr Lys Ala Leu Leu Ala Tyr Ala Phe Ala Leu Ala Gly
1155 1160 1165

Asn Gln Asp Lys Arg Lys Glu Val Leu Lys Ser Leu Asn Glu Glu Ala
1170 1175 1180

Val Lys Lys Asp Asn Ser Val His Trp Glu Arg Pro Gln Lys Pro Lys
1185 1190 1195 1200

Ala Pro Val Gly His Phe Tyr Glu Pro Gln Ala Pro Ser Ala Glu Val
1205 1210 1215

Glu Met Thr Ser Tyr Val Leu Leu Ala Tyr Leu Thr Ala Gln Pro Ala
1220 1225 1230

Pro Thr Ser Glu Asp Leu Thr Ser Ala Thr Asn Ile Val Lys Trp Ile
1235 1240 1245

Thr Lys Gln Gln Asn Ala Gln Gly Gly Phe Ser Ser Thr Gln His Thr
1250 1255 1260

Val Val Ala Leu His Ala Leu Ser Lys Tyr Gly Ala Ala Thr Phe Thr
1265 1270 1275 1280

Arg Thr Gly Lys Ala Ala Gln Val Thr Ile Gln Ser Ser Gly Thr Phe
1285 1290 1295

Ser Ser Lys Phe Gln Val Asp Asn Asn Asn Arg Leu Leu Leu Gln Gln
1300 1305 1310

Val Ser Leu Pro Glu Leu Pro Gly Glu Tyr Ser Met Lys Val Thr Gly
1315 1320 1325

Glu Gly Cys Val Tyr Leu Gln Thr Ser Leu Lys Tyr Asn Ile Leu Pro
1330 1335 1340

Glu Lys Glu Glu Phe Pro Phe Ala Leu Gly Val Gln Thr Leu Pro Gln
1345 1350 1355 1360

Thr Cys Asp Glu Pro Lys Ala His Thr Ser Phe Gln Ile Ser Leu Ser
1365 1370 1375

Val Ser Tyr Thr Gly Ser Arg Ser Ala Ser Asn Met Ala Ile Val Asp
1380 1385 1390

Val Lys Met Val Ser Gly Phe Ile Pro Leu Lys Pro Thr Val Lys Met
1395 1400 1405

Leu Glu Arg Ser Asn His Val Ser Arg Thr Glu Val Ser Ser Asn His
1410 1415 1420

Val Leu Ile Tyr Leu Asp Lys Val Ser Asn Gln Thr Leu Ser Leu Phe
1425 1430 1435 1440

Phe Thr Val Leu Gln Asp Val Pro Val Arg Asp Leu Lys Pro Ala Ile
1445 1450 1455

Val Lys Val Tyr Asp Tyr Tyr Glu Thr Asp Glu Phe Ala Ile Ala Glu
1460 1465 1470

Tyr Asn Ala Pro Cys Ser Lys Asp Leu Gly Asn Ala
1475 1480

PATENT CLAIMS

1. A process for the production of recombinant α-macroglobulin, variants, fragments or derivatives thereof, wherein a functionally operative expression vector comprising a gene encoding for the expression of α-macroglobulin, variants, fragments or derivatives thereof, or alleles of such a gene, is introduced into a suitable host capable of expressing said gene, said host is cultured in a suitable nutrient medium containing sources of assimilable carbon and nitrogen and other essential nutrients, and the expressed α-macroglobulin or fragments or derivatives thereof is recovered.

2. The process of claim 1, wherein said gene encodes for the expression of human α_2-macroglobulin, variants, fragments or derivatives thereof.

3. The process of claim 2, wherein said gene encodes for the expression of human α_2-macroglobulin having the amino acid sequence of SEQ ID NO:2, or a fragment or derivative thereof.

4. The process of claim 2 or 3, wherein said gene comprises the DNA sequence of SEQ ID NO:1, or a fragment thereof.

5. The process of claim 1 or 2, wherein said gene encodes for a variant α-macroglobulin, in which the amino acid sequence of the bait region has been altered.

6. The process of claim 5, wherein the bait region has been altered by incorporation of further proteinase target sites.

7. The process of claim 5, wherein the bait region has been altered by removal of proteinase target sites.

8. The process of claim 5, wherein the bait region has been altered by replacing one or more specific proteinase target sites with one or more other specific proteinase target sites.

9. The process of claim 8, wherein said proteinase target sites are specific for bovine trypsin, Streptomyces griseus trypsin, papain, porcine elastase, bovine chymosin, bovine chymotrypsin, Staphylococcus aureus strain

V8 proteinase, human plasmin, bovine thrombin, thermolysin, subtilisin Novo and/or Streptomyces griseus proteinase B.

10. The process of claim 5, wherein wherein the bait region has been altered by replacing said bait region or part thereof with a bait region or a part thereof from another α-macroglobulin.

11. The process of claim 10, wherein said bait regions originate from human α_2M, Pregnancy Zone Protein (PZP), rat α_1M, rat α_2M, rat α_1I_3 variant 1, or rat α_1I_3 variant 2 (α_1I_3 = α_1-inhibitor 3), especially PZP.

12. The process of any of claims 5 to 11, wherein said gene encodes for the expression of human a α_2-macroglobulin variant having the amino acid sequence of SEQ ID NO:4, or a fragment or derivative thereof.

13. The process of any of claims 5 to 12, wherein said gene comprises the DNA sequence of SEQ ID NO:3, or a fragment thereof.

14. The process of any of the claims 1 to 13, wherein said gene is a synthetic gene.

15. The process of any of the claims 1 to 14, wherein said α-macroglobulin, variant, fragment or derivative thereof is co-expressed with a desired gene product.

16. The process of any of the claims 1 to 15, wherein said gene is, or is derived from, a human gene.

17. The process of any of the claims 1 to 16, wherein said host is a bacterial strain, a fungal strain, a mammalian cell line, or a mammal.

18. The process of claim 17, wherein said host is a fungus.

19. The process of claim 18, wherein said fungus belongs to the genus Aspergillus.

20 The process of claim 18, wherein said host is a yeast.

21. The process of claim 20, wherein said yeast belongs to the genus Saccharomyces.

22. The process of claim 17, wherein said host is a mammalian cell line.

23. The process of claim 22, wherein said mammalian cell line is a Syrian Baby Hamster Kidney (BKH) cell line.

24. The process of claim 23, wherein said cell line is available from ATCC under No. CRL 1632.

25. A DNA sequence comprising a gene encoding for the expression of an α-macroglobulin, variants, fragments or derivatives thereof.

26 The DNA sequence of claim 25, wherein said gene encodes for human α_2-macroglobulin.

27. The DNA sequence of claim 25, wherein said gene encodes for the amino acid sequence of SEQ ID NO:2 or a fragment or derivative thereof.

28. The DNA sequence of claim 26 or 27, wherein said gene has the nucleotide sequence of SEQ ID NO:1 or a fragment thereof.

29. The DNA sequence of claim 25 or 26, wherein said gene encodes for a variant α-macroglobulin, in which the amino acid sequence of the bait region has been altered.

30. The DNA sequence of claim 29, wherein said bait region has been altered by incorporation of further proteinase target sites.

31. The DNA sequence of claim 29, wherein said bait region has been altered by removal of proteinase target sites.

32. The DNA sequence of claim 29, wherein said bait region has been altered by replacing one or more specific proteinase target sites with one or more other specific proteinase target sites.

33. The DNA sequence of claim 29, wherein, wherein said proteinase target sites are specific for bovine trypsin, Streptomyces griseus trypsin, papain, porcine elastase, bovine chymosin, bovine chymotrypsin, Staphylococcus aureus strain V8 proteinase, human plasmin, bovine thrombin, thermolysin, subtilisin Novo and/or Streptomyces griseus proteinase B.

34. The DNA sequence of claim 29, wherein the bait region has been altered by replacing said bait region or part thereof with a bait region or a part thereof from another α-macroglobulin.

35. The DNA sequence of claim 34, wherein said bait region originates from human α_2M, Pregnancy Zone Protein (PZP), rat α_1M, rat α_2M, rat α_1I_3 variant 1, or rat α_1I_3 variant 2, especially PZP.

36. A functionally operative expression vector comprising a gene in accordance with any of the claims 25 to 35 for the expression of human α_2-macroglobulin, variants, fragments or derivatives thereof, or alleles of such a gene.

37. The vector of claim 36, further comprising regulatory elements necessary for the stable maintenance of said vector in mammalian cells.

38. The vector of claim 36 or 37, further comprising sequences providing for the processing and secretion of the expressed product.

39. The vector of any of the claims 36 to 38, further comprising one or more other genes encoding for a desired gene product.

40. A functionally operative expression vector comprising a gene encoding for the expression of an α-macroglobulin, variants, fragments or derivatives thereof, or alleles of such a gene, essentially as described.

41. A transformed host comprising a functionally operative expression vector comprising a gene encoding for the expression of human α_2-macroglobulin or fragments or derivatives thereof, or alleles of such a gene.

42. The host of claim 41, wherein said vector is the vector of any of the claims 36 to 40.

43. The host of claim 41 or 42, wherein said host is a bacterial strain, a fungal strain, a mammalian cell line, or a mammal.

44. The host of claim 43, wherein said host is a fungus.

45. The host of claim 44, wherein said fungus belongs to the genus Aspergillus.

46. The host of claim 44, wherein said host is a yeast.

47. The host of claim 46, wherein said host belongs to the genus Saccharomyces.

48. The host of claim 43, wherein said host is a mammalian cell line.

49. The host of claim 48, wherein said host is a Syrian Baby Hamster Kidney (BHK) cell line.

50. The host of claim 49, wherein said cell line is available from ATCC under No. CRL 1632.

51. Recombinant human α_2-macroglobulin of SEQ ID NO:2 or SEQ ID NO:4 in an active form.

52. Recombinant α-macroglobulin, variants, fragments or derivatives thereof produced by a process of any of the claims 1 to 24.

53. Recombinant α-macroglobulin, variants, fragments or derivatives thereof of claim 52 produced by the use of a vector of any of the claims 36 to 40.

54. Recombinant α-macroglobulin, variants, fragments or derivatives thereof essentially as described.

55. Recombinant human α_2-macroglobulin, variants, fragments or derivatives thereof essentially as described.

56. A growth medium comprising one or more α-macroglobulins.

57. A growth medium comprising recombinant α-macroglobulin, variants, fragments or derivatives thereof according to any of the claims 51 to 55.

58. Use of recombinant α-macroglobulin, variants, fragments or derivatives thereof according to any of the claims 51 to 55 as a protein carrier in enzyme replacement therapy.

59. Use of recombinant α-macroglobulin, variants, fragments or derivatives thereof according to any of the claims 51 to 55 as a DNA carrier in gene therapy.

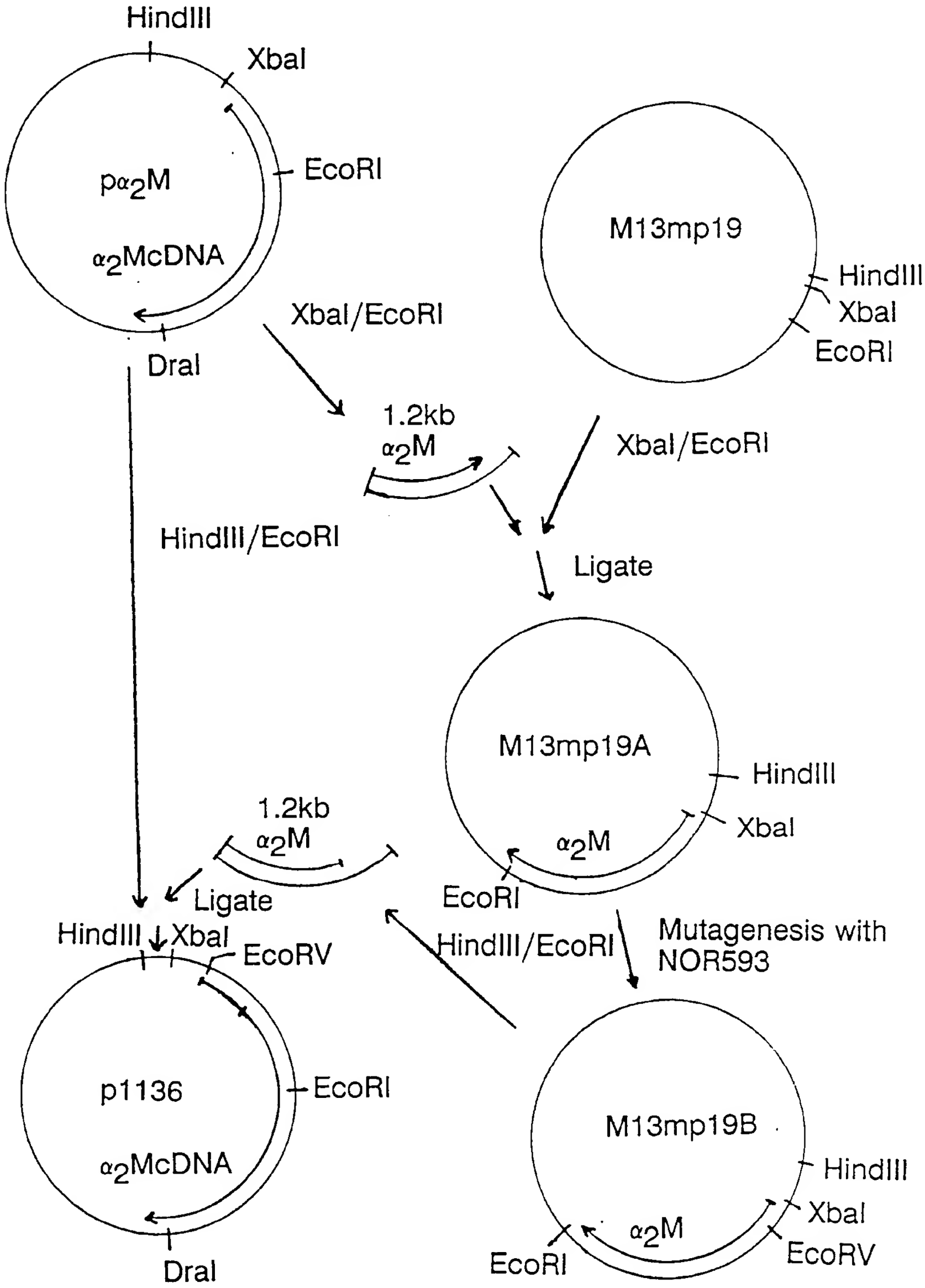
HindIII
XbaI
EcoRI
pα2M
α2McDNA
DraI
XbaI/EcoRI
1.2kb
α2M
HindIII/EcoRI
M13mp19
HindIII
XbaI
EcoRI
XbaI/EcoRI
Ligate
M13mp19A
HindIII
XbaI
α2M
EcoRI
1.2kb
α2M
Ligate
HindIII
XbaI
EcoRV
HindIII/EcoRI
Mutagenesis with
NOR593
p1136
EcoRI
α2McDNA
DraI
M13mp19B
HindIII
XbaI
α2M
EcoRI
EcoRV

FIG. 1A

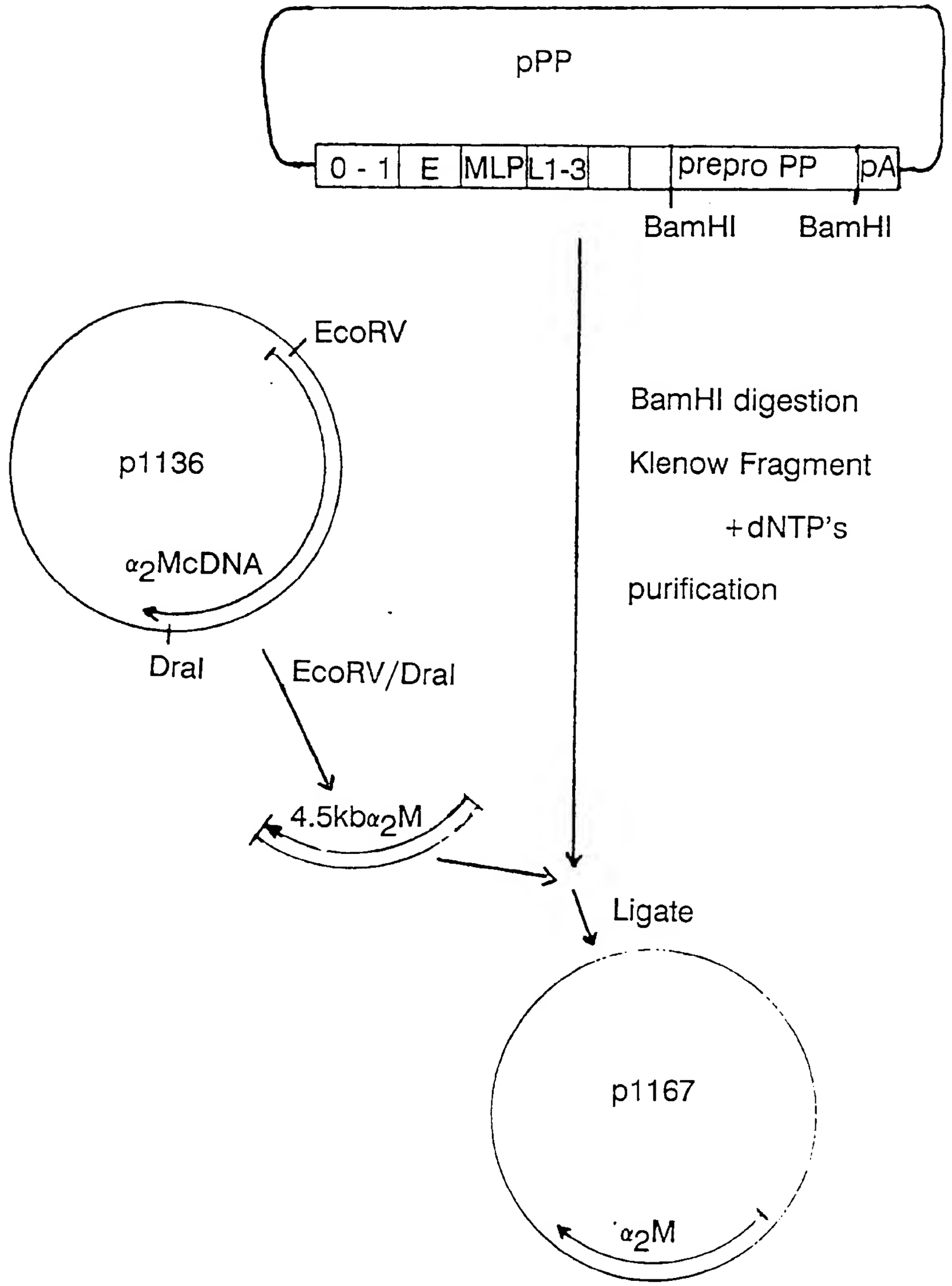
pPP
0 - 1
E
MLP
L1-3
prepro PP
pA
BamHI
BamHI
BamHI digestion
Klenow Fragment
+dNTP's
purification
EcoRV
p1136
α2McDNA
DraI
EcoRV/DraI
4.5kbα2M
Ligate
p1167
α2M

FIG. 1B

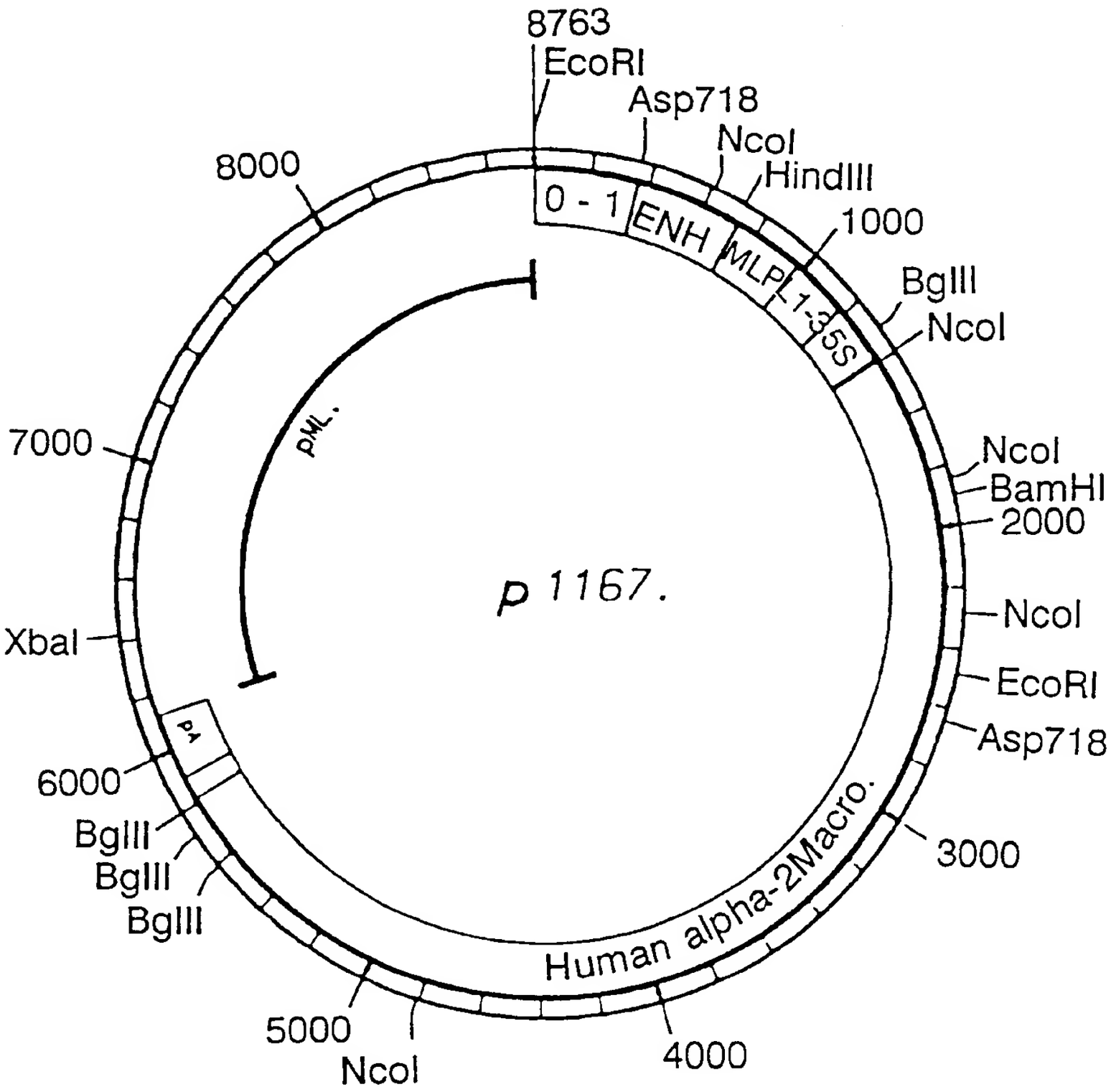
8763
EcoRI
Asp718
NcoI
HindIII
1000
BglII
NcoI
NcoI
BamHI
2000
NcoI
EcoRI
Asp718
3000
4000
Ncol
5000
BglII
BglII
BglII
6000
XbaI
7000
8000
0 - 1
ENH
MLP
1-35S
pML.
p 1167.
pA
Human alpha-2Macro.

FIG. 2

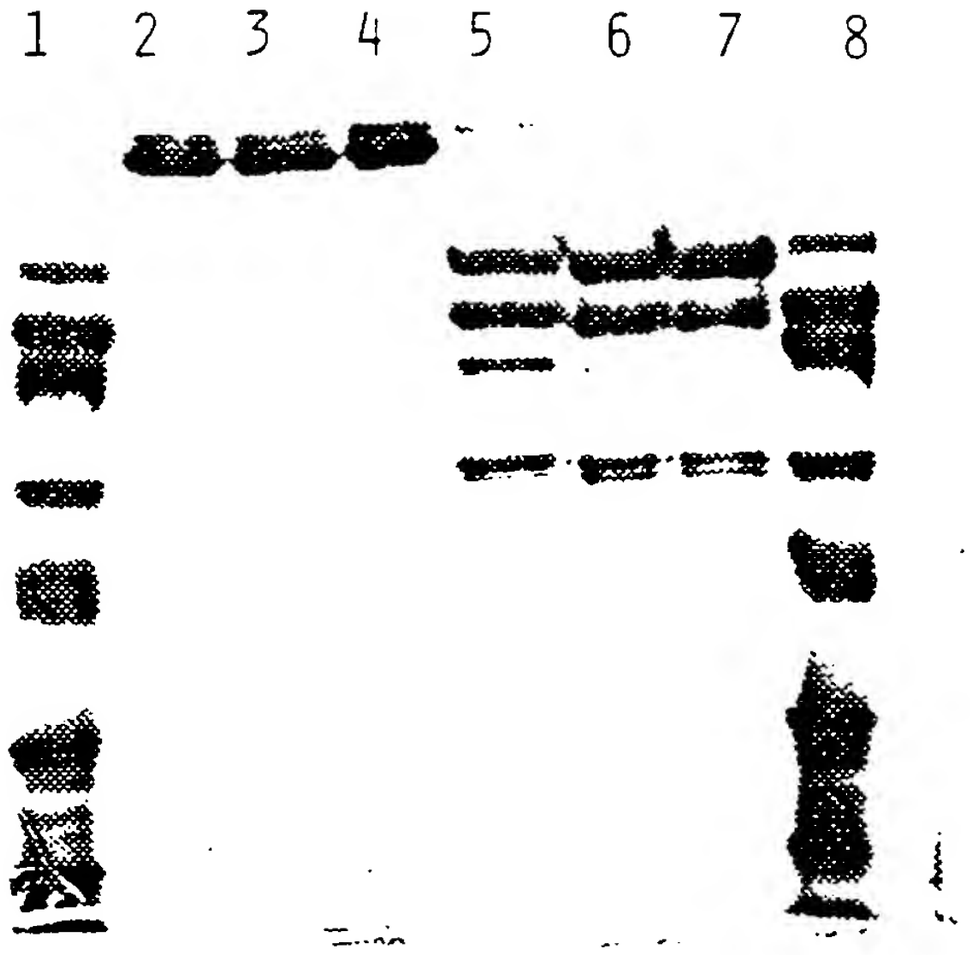
1 2 3 4 5 6 7 8
NATIVE 10-20%
360 kD
180 kD
120 kD
60 kD

FIG. 3

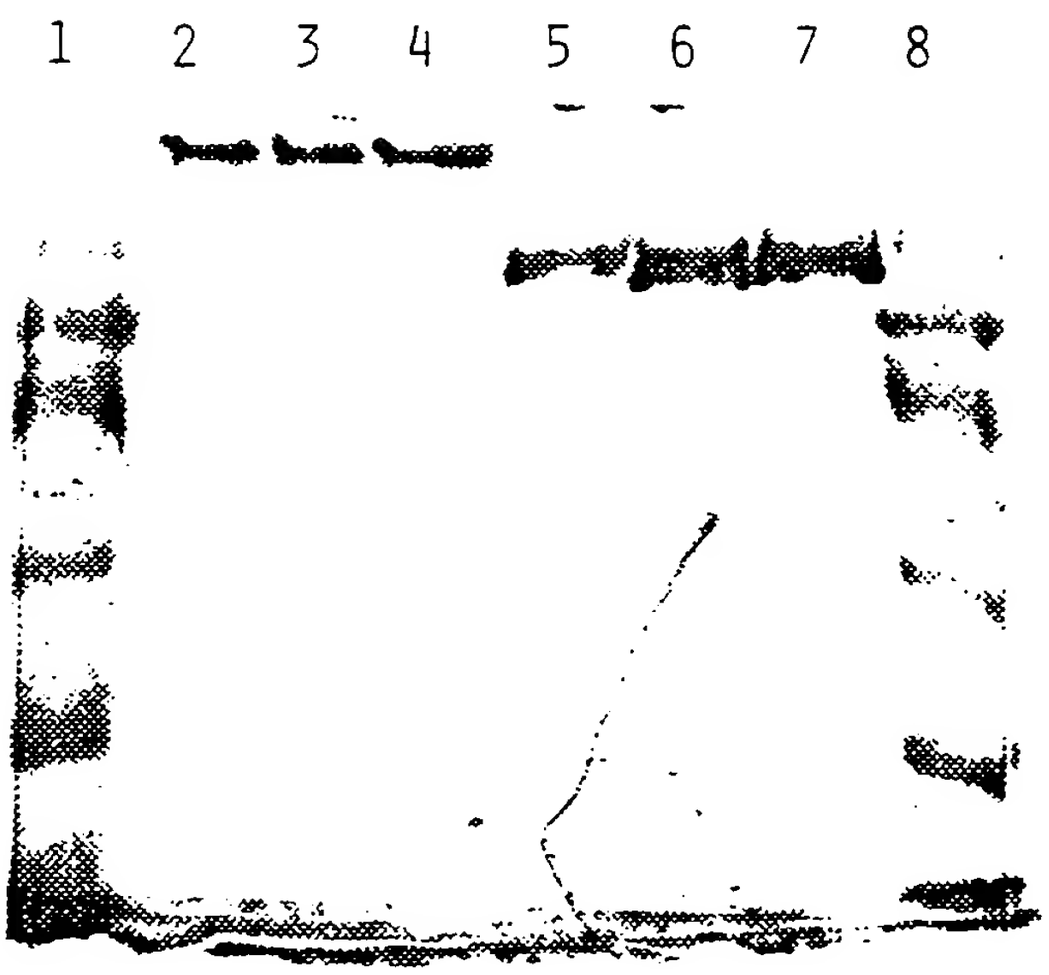
1 2 3 4 5 6 7 8
360 kD
180 kD

FIG. 4

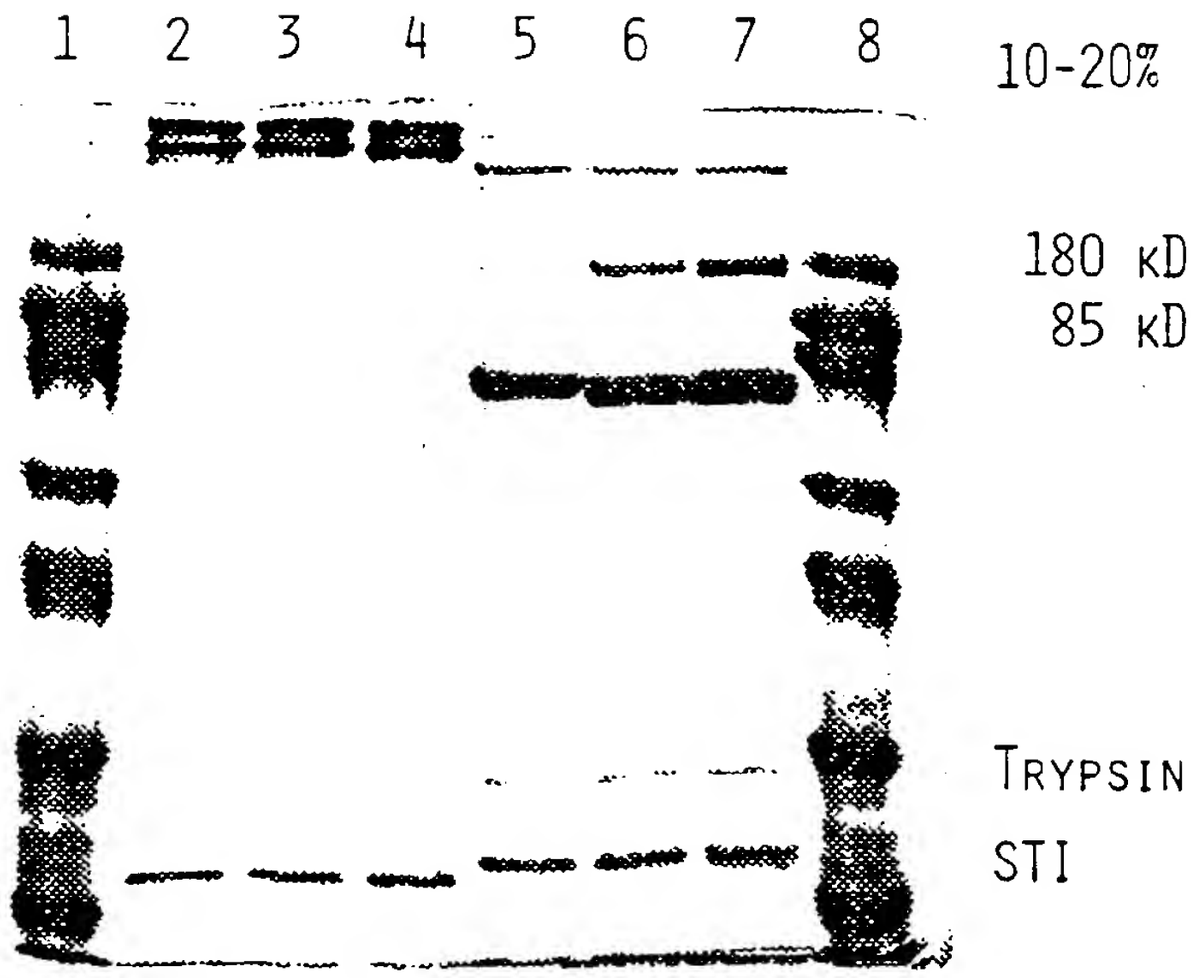
1 2 3 4 5 6 7 8
10-20%
180 kD
85 kD
Trypsin
STI

FIG. 5

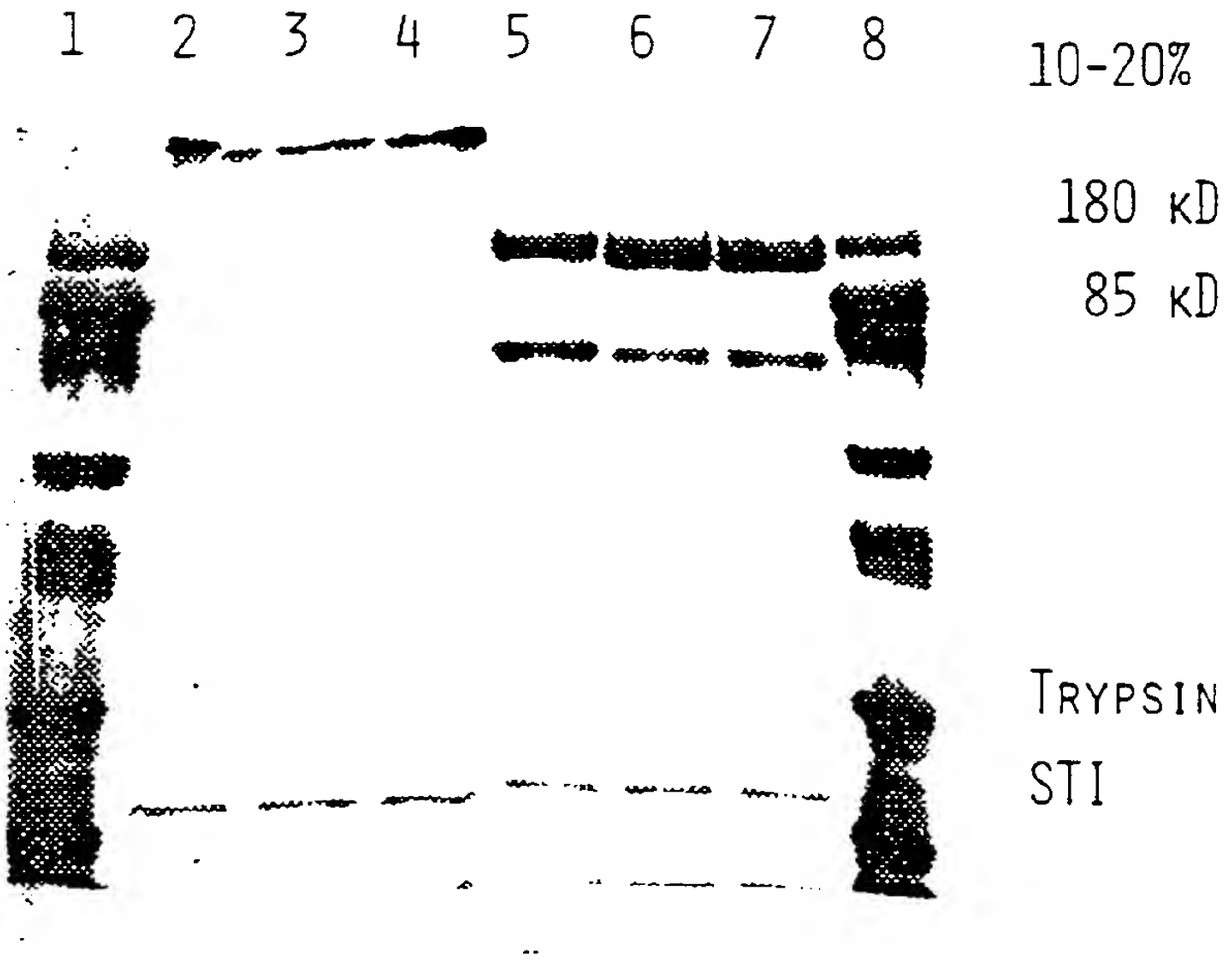
1 2 3 4 5 6 7 8
10-20%
180 kD
85 kD
Trypsin
STI

FIG. 6

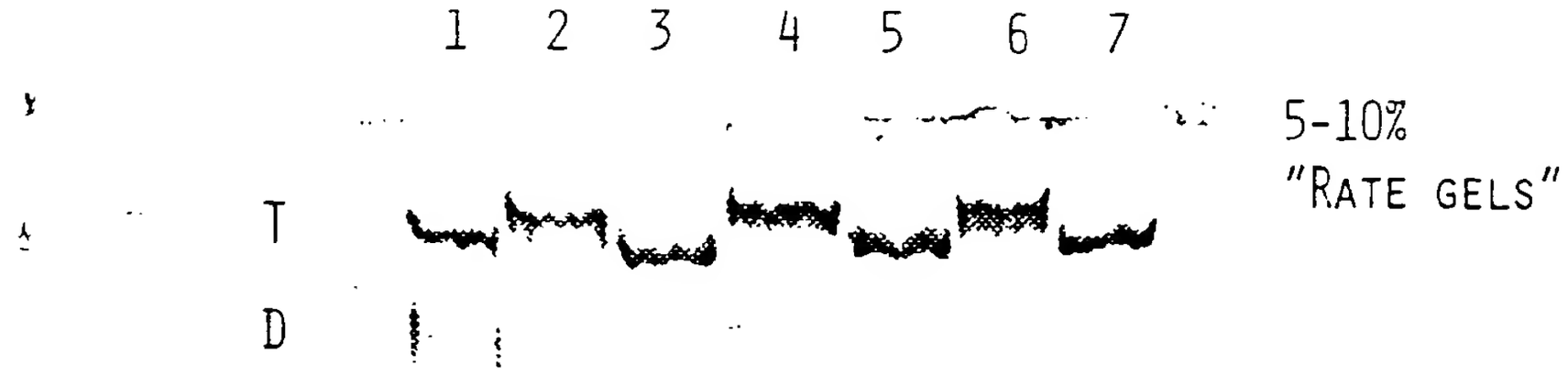

FIG. 7

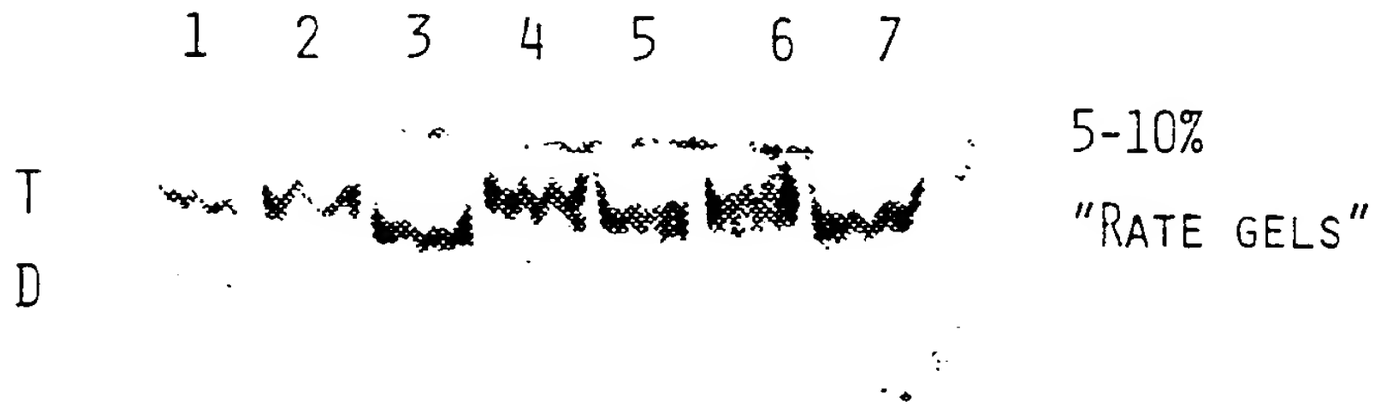

FIG. 8

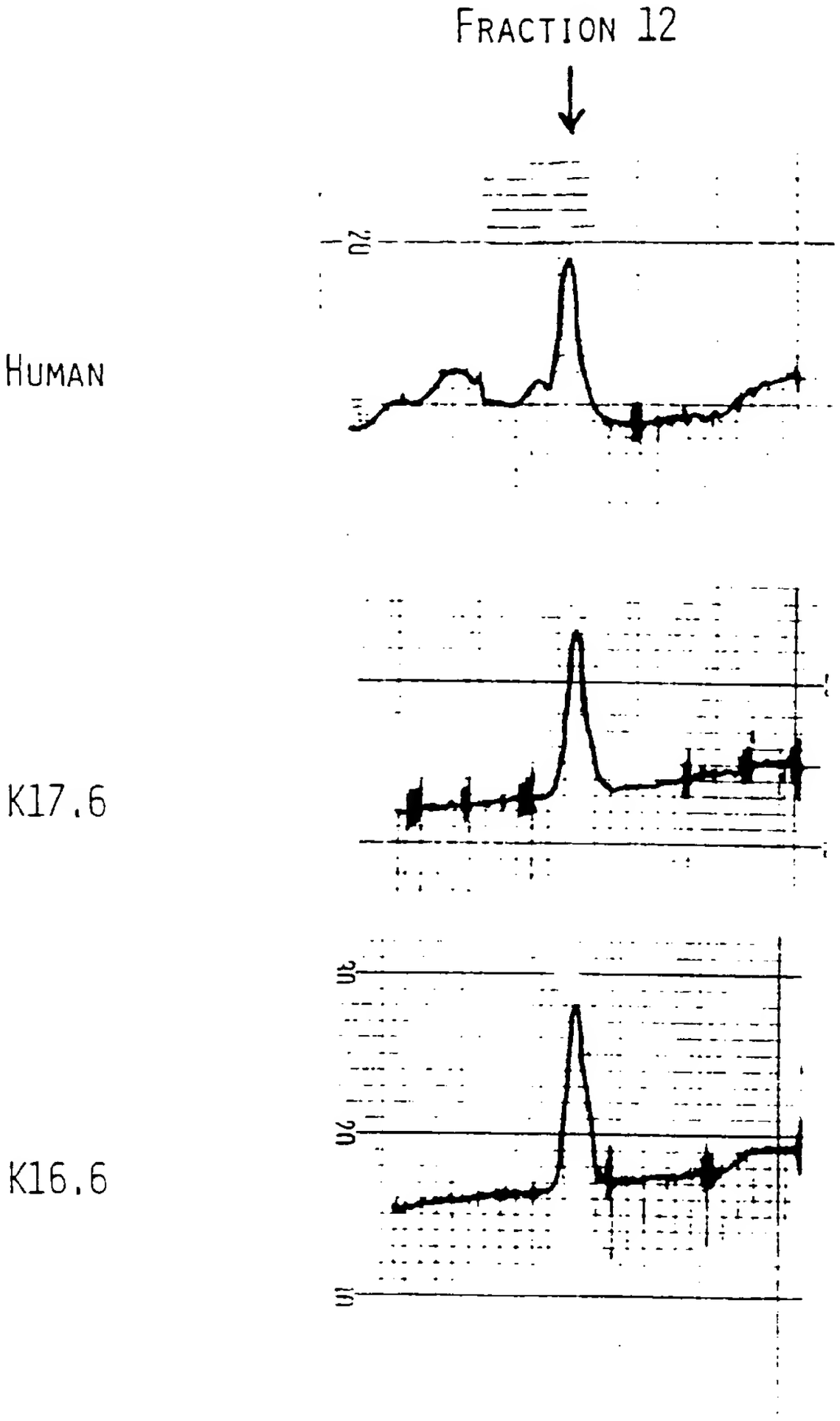

FIG. 9

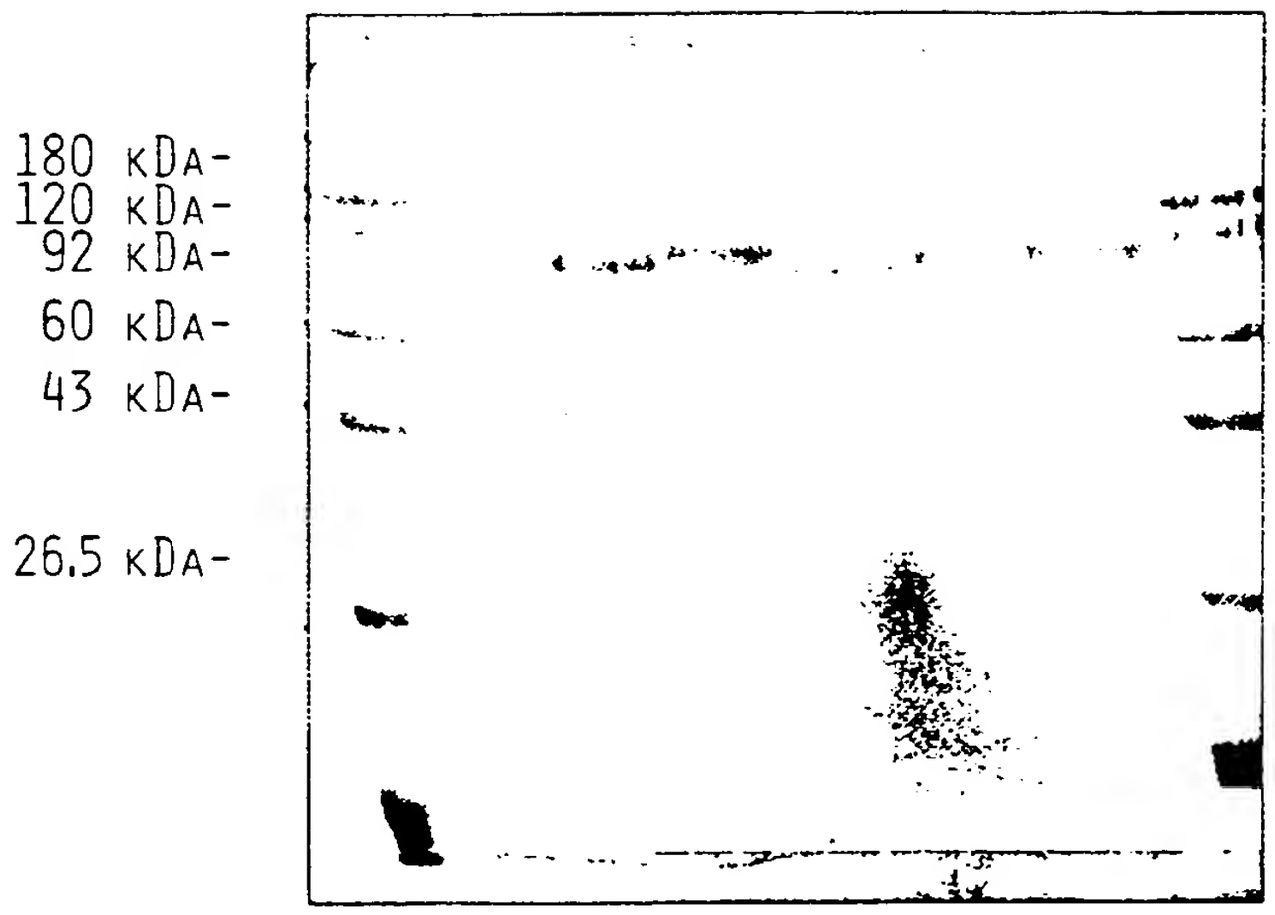
1 2 3 4 5 6 7 8
180 kDa-
120 kDa-
92 kDa-
60 kDa-
43 kDa-
26.5 kDa-
Reducing
10-20%

FIG. 10

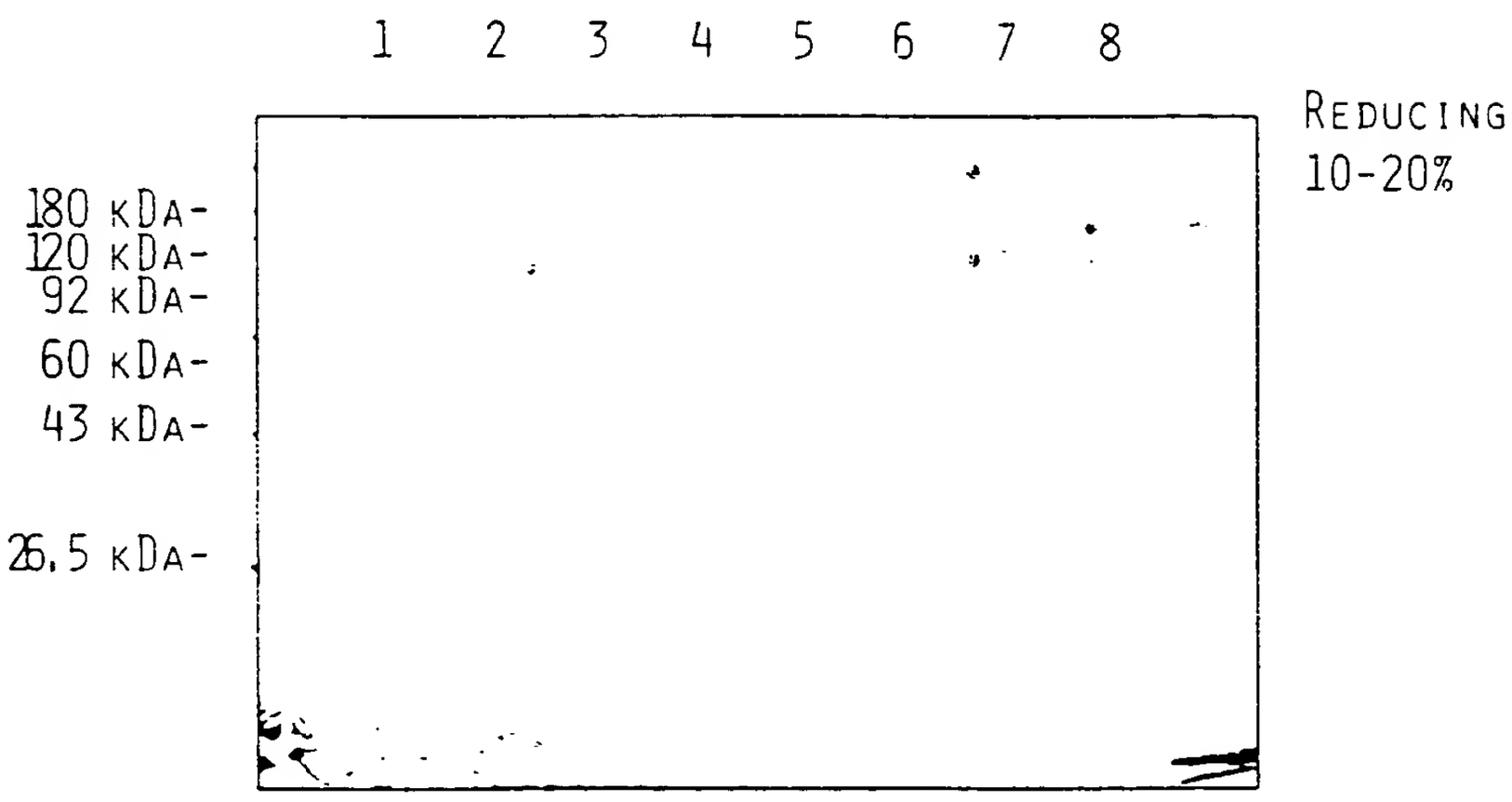
1 2 3 4 5 6 7 8
180 kDa-
120 kDa-
92 kDa-
60 kDa-
43 kDa-
26.5 kDa-
Reducing
10-20%

FIG. 11

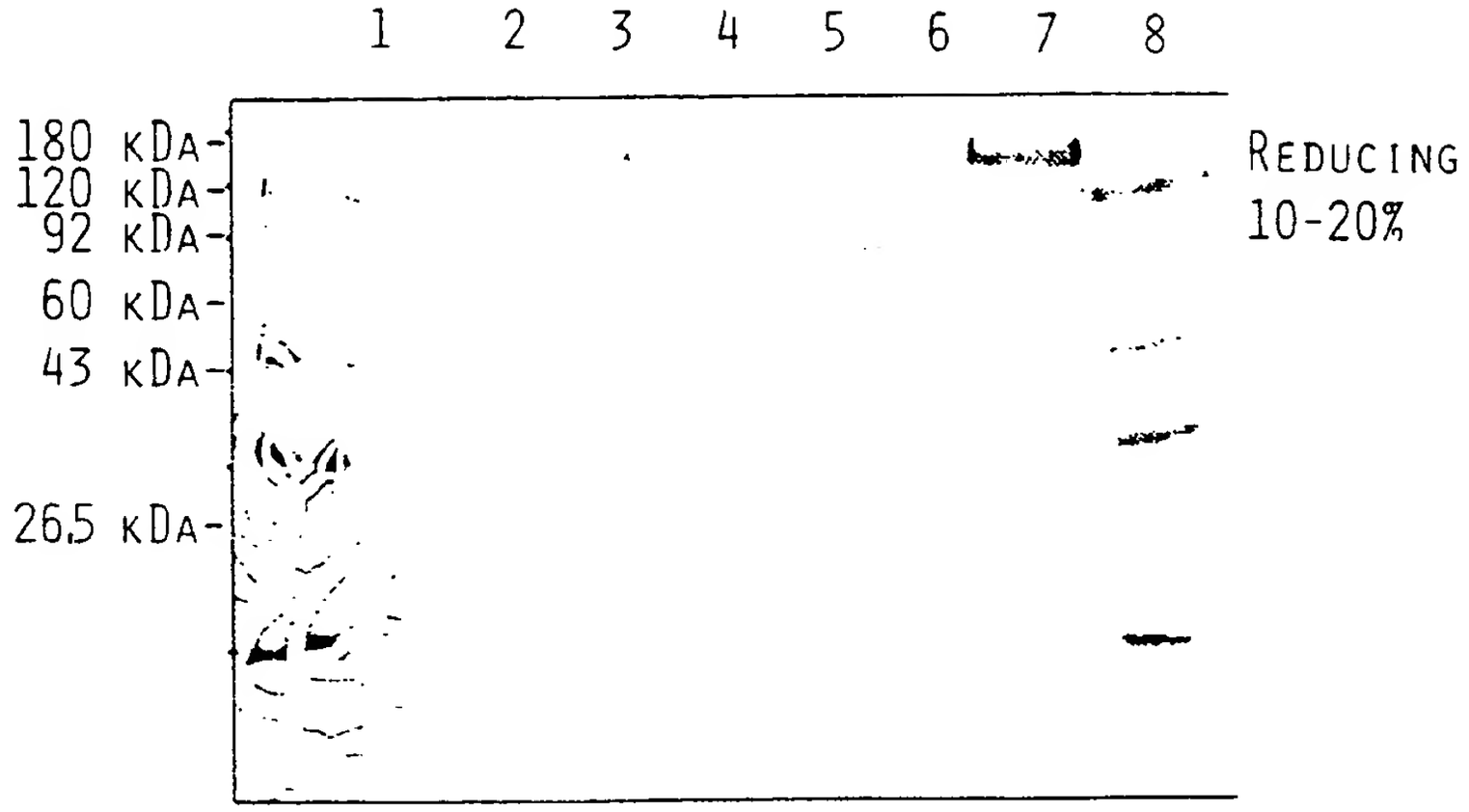
1 2 3 4 5 6 7 8
180 kDa-
120 kDa-
92 kDa-
60 kDa-
43 kDa-
26.5 kDa-
Reducing
10-20%

FIG. 12

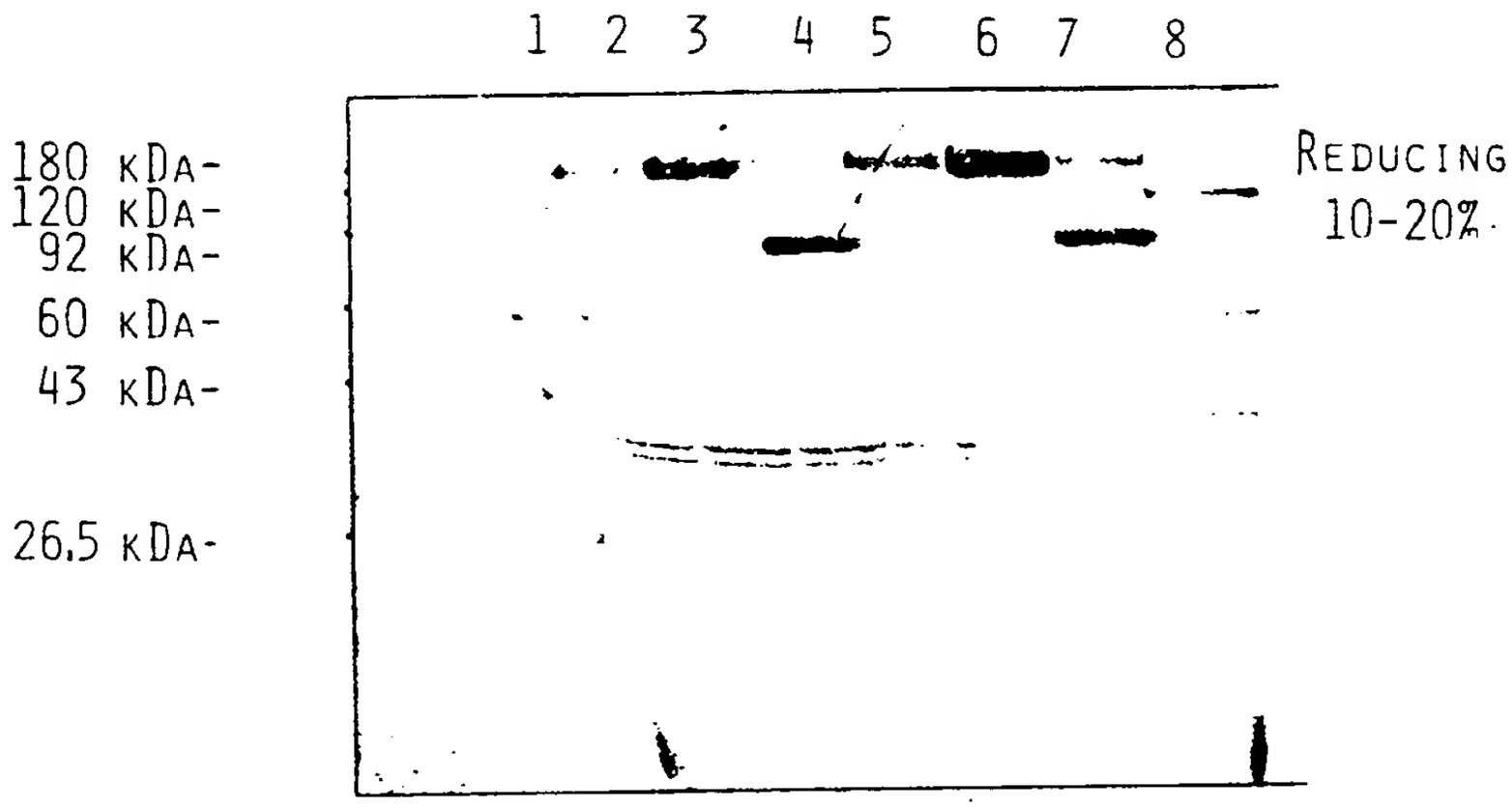
1 2 3 4 5 6 7 8
180 kDa-
120 kDa-
92 kDa-
60 kDa-
43 kDa-
26.5 kDa-
Reducing
10-20%

FIG. 13

INTERNATIONAL SEARCH REPORT

International Application No PCT/DK 90/00225

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC5: C 12 N 15/15, A 61 K 37/64, C 07 K 13/00, C 12 P 21/02

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
IPC5	A 61 K; C 12 N; C 07 K

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in Fields Searched[8]

SE,DK,FI,NO classes as above

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category *	Citation of Document,[11] with indication, where appropriate, of the relevant passages[12]	Relevant to Claim No.[13]
X	Proc Natl Acad Sci USA, Vol. 82, No. 8, 1985 Kan, Chen Chen et al.: "Nucleotide sequence of cDNA encoding human alpha-2-macroglobulin and assignment of the chromosomal locus", see page 2282 - page 2286	1,2,4, 14-28, 36-50, 52-59
Y	--	5-10,29-34
Y	Chemical Abstracts, volume 96, no. 15, 12 April 1982, (Columbus, Ohio, US), Mortensen, steen Bet al.: "Primary and secondary cleavage sites in the bait region of alpha-2-macroglobulin ", see page 253, abstract 117756z, & FEBS Lett 1981, 135(2), 295-300	5-10,29-34
A	--	1-59
	Chemical Abstracts, volume 95, no. 7, 17 August 1981, (Columbus, Ohio, US), Sottrup-Jensen, Lars et al.: "Primary structure of the'bait' region for proteinases in alpha-2-macroglobulin. Nature of the complex ", see page 261, abstract 57059s, & FEBS Lett 1981, 127(2), 167-173	
	--	

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
11th December 1990	1990 -12- 13
International Searching Authority	**Signature of Authorized Officer**
SWEDISH PATENT OFFICE	Yvonne Siösteen

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
A	Chemical Abstracts, volume 101, no. 11, 10 September 1984, (Columbus, Ohio, US), Sottrup-Jensen, Lars et al.: "Primary structure of human alpha-2-macroglobulin. V. The complete structure ", see page 237, abstract 85952p, & J. Biol. Chem. 1984, 259(13), 8318-8327 --	1-59
P	Chemical Abstracts, volume 111, no. 23, 4 December 1989, (Columbus, Ohio, US), Sottrup-Jensen, Larset al.: "The alpha-macroglobulin bait region. Sequence diversity and localization of cleavage sites for proteinases in five mammalian alpha-macroglobulins ", see page 227, abstract 210722y, & J. Biol. chem. 1989, 264(27), 15781-15789 --	5-10,29-34
P	Chemical Abstracts, volume 112, no. 25, 18 June 1990, (Columbus, Ohio, US), Marynen, P et al.: "A genetic polymorphism in a functional domain of human pregnancy zone protein: the bait region. Genomic structure of the bait domains of human pregnancy zone protein and alpha-2-macroglobulin ", see page 167, abstract 230679p, & FEBS Lett. 1990, 262(2), 349-352 --	5-10,29, 34

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. PCT/DK 90/00225

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the Swedish Patent Office EDP file on 90-11-01
The Swedish Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date